Raytheon

Air Traffic Management System Development and Integration (ATMSDI)

CTOD-2.3-2
Draft Guidelines
Subtask 4 – Human Factors Metrics Guidelines

December 2001

Contract No.: NAS2-00015

Prepared for

NASA Ames Research Center
Richard Mogford, Ph.D.

Prepared by

Raytheon ATMSDI Team
Nicole Casso, Parimal Kopardekar, Ph.D. (Titan Systems)

ACKNOWLEDGEMENT

The authors sincerely thank Dr. Richard Mogford for providing ideas and literature, and encouraging discussions.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONT.)

LIST OF ILLUSTRATIONS

LIST OF TABLES

EXECUTIVE SUMMARY

The benefits and feasibility assessment of distributed air-ground traffic management (DAG-TM) is an important consideration before the Concept Elements (CEs) are fully developed. Human factors metrics provide insight into how the CEs could affect the human performance and hence the system performance. Therefore, it is important to study the human factors benefits and feasibility metrics at each Technology Readiness Level (TRL).

This guidelines document and literature review is a compilation of known and needed human factors metrics applicable to the air traffic management environment. This database of metrics will be used to develop guidelines for a battery of metrics for each CE across the TRLs to be used in all DAG-TM evaluations.

This Page Intentionally Left Blank

1. BACKGROUND

Distributed air-ground traffic management (DAG-TM) represents a paradigm shift that will bring significant changes to the roles and responsibilities of air traffic controller providers (ATSP), traffic flow management specialists (TFMS), flight crew (FC), and airline operations center specialists (AOCS). These new roles and responsibilities will require different decision support tools and procedures.

Since these changes will alter the human tasks and allocation of man/man and man/machine functions, the benefits and feasibility of DAG-TM concepts must be investigated before matured concepts, procedures, and decision support tools are developed. This document presents guidelines in the use of human factors metrics to determine the feasibility and benefits of DAG-TM concepts.

2. SCOPE

The focus of this guidelines document is primarily on the human factors metrics related to the air traffic management domain. Guidelines are drawn from the set of metrics found in the literature review (see Appendix) and narrowed down to a select battery of metrics to be used across DAG-TM concept element (CE) 5 Free Maneuvering and CE 11 Self-spacing for In-trail and Merging.

3. OBJECTIVE

The objective of this guidance document is to develop a suggested standard toolset of metrics to be used in future DAG-TM studies. Included in this document is a literature review (see appendix) intended to provide an extensive summary database of current human performance metrics. In addition to recommended metrics, the purpose of this document is to identify areas that require the development of new metrics for future concepts.

4. LITERATURE REVIEW

Researchers conducted a literature review of performance metrics in the air traffic domain. The table presented in the Appendix is a compilation of human factors and system metrics collected in a range of human factors studies. Several authors have performed reviews of various metrics (Buckley, DeBaryshe, Hitchner, & Kohn 1983; Hadley, Guttman, & Stringer, 1999; Gawron, 2000). As these documents stress the air traffic control aspects of performance measures, this literature review document serves as an extension of these previous efforts. It is necessary to also identify metrics pertinent to the flight deck and airline operations center, as well as identify metrics needed for the distributed management concept. It is recognized that this review may not be complete, and is intended to be a living document, with metrics added as they are identified or developed.

A detailed description of CE 5 can be found in Philips (2000). The potential benefits of CE 5 operations are as follows:

- Reduction in excessive and non-preferred deviations for separation assurance and local TFM conformance, due to the ability of the flight crew (for equipped aircraft) to self-separate and maintain local TFM conformance according to their preferences.
- Increased safety in separation assurance for all aircraft, due to communications, navigation, and surveillance redundancy (FC as primary and ATC as backup) and increased situational awareness of the FC of appropriately equipped aircraft.
- Reduced ATSP workload for separation assurance and local TFM conformance plus reduced FC workload for communications, due to the distribution of responsibility for separation assurance and local TFM conformance between the ATSP and appropriately equipped FCs.

A detailed description of CE 11 can be found in Sorensen (2000).

- Increased arrival capacity/throughput in IMC, due to a reduction in excessive spacing buffers resulting from the ability of appropriately equipped aircraft to operate as if they were in VMC.
- Reduced ATSP workload, due to transfer of separation responsibility to the flight crew of appropriately equipped aircraft.

The metrics described in this guidelines document are specifically selected such that they will be able to assess the potential benefits of DAG-TM CEs.

5. METRIC GUIDELINES OVERVIEW

Researchers used the literature review database (see Appendix) to derive guidelines for metrics for CE 5 and 11, and to determine the appropriate metrics that will assess the benefits and feasibility of each CE across metric constructs. Guidelines generally follow database constructs (workload, situation awareness, complexity, etc.), and present a selection of metrics for each. When a metric is dependent on a specific tool or a technique, details such as description, system requirements, strengths and limitations, acquisition information, and relevant studies involving each metric will be presented in a tabular format (see Table 1). In this example, Air Traffic Workload Input Technique (ATWIT) is a suggested measure for workload. Other recommendations will be made for workload measures, and this method will follow for other constructs where applicable. More general metrics that do no rely on a specific modality (e.g., delays) will be presented within each category (e.g., capacity).

Table 1. Sample Metric Guideline.

Metric Construct:	Workload.
Metric:	Air Traffic Workload Input Technique (ATWIT).
Description:	Subjective workload measured at standard intervals during the simulation. Array presents up to 10 input buttons. Each key corresponds to a workload rating. At predetermined intervals, low-level beep is emitted and lights of a numbered Workload Assessment Keyboard (WAK) are illuminated for a specified time interval (i.e., 20 seconds), after which, participants are instructed to enter a subjective workload rating. If the participant is unable to enter a rating by the end of the allotted time, an automatic rating of 99 is recorded.
Requirements:	Workload Assessment Keyboard (WAK) or other functional equivalent input device.
Strengths:	Reliable. Real-time ratings.
Limitations:	Minimal disruption of task.
Implementation:	Real time. Rating interval is variable. Portable. Instantaneous data extraction.
Acquisition Information:	A & J Industries, (405) 794-6667. POC: Alie Burgin.
Estimated Cost:	$3,000.00 (four WAK units and multiplexer). Programming is required.
References:	FAA & NASA 2001; Porterfield, 1997; Sollenberger, Stein & Gromelski, 1997.

6. GUIDELINES

The intent of the guidelines document is to create a selected set of metrics that may be used throughout DAG-TM research studies. The hope is that by standardizing the measures, there will be consistency for comparisons across studies. These metrics were selected based on the literature review database (see appendix), which is composed of a vast amount of system and human performance metrics. The literature review revealed many performance metrics for the ATSP and FC. Although some of the general metrics may be applied to the AOC, further research is needed for guidelines in this area. Some metrics from the literature review database are used in specific applications that may not be relevant or useful for DAG-TM applications. Other measures are redundant and are therefore excluded. Table 2 is a summary of recommended metrics for DAG-TM research.

Table 2. Metric Constructs

Constructs	Metrics
Capacity	Throughput; Aircraft under control; Time in sector; Average separation-en route-terminal; Time between arrivals; Frequency utilization; Spacing on final; and Airport capacity usage: gate, runway, taxiway, arrival, and departure.
Efficiency	Available altitude, Fuel burn, Delay, Optimal speed assignments, Delta preferred path, and CPA above standard separation.
Flexibility	Available altitudes, Degrees of freedom, Weather deviations, Trajectory to optimize local flow constraints, and Exclusionary vs. mixed equipage.
Safety	Operational errors; Operational deviations; Other errors (keyboard entry, hear-back/read-back etc); Conflict alerts: number, duration, and type; and CPA below standard separation.
Workload	Communications: duration, frequency, type, and errors; ATWIT; POSWAT; NASA TLX; Control input: FC and ATSP keyboard entries, FD maintenance; Data block manipulation/overlap; and Flight control input.
Decision Making	Time to make decision, Time to share information, Impact on efficiency, and Impact on workload.
Situation Awareness	SAGAT, SART, Missed handoffs, Detection of conflict before automated alert, and Halo Initiation.
Usability	Subjective questionnaire, Number of incorrect actions, Number of times “help” was accessed, Information presentation and interaction, and Learning curve.
Complexity	DD variables.
Simulation Fidelity	Traffic characteristics, Other simulation characteristics, Realism rating, Impact of technical problems, Impact of pseudopilots, and Scenario difficulty.
Other	Trust in automation, Trust in other agents.

6.1 CAPACITY

6.1.1 Background

Increasing traffic and delays have perpetuated the thrust of air traffic research focused on capacity. Utilization of system capacity related to air traffic operations is captured by aircraft counts, throughput, arrivals, time in sector, and delays (Allendoerfer & Galushka, 1999; FAA & NASA 2001; Magyarits & Kopardekar, 2000). System tools and procedures measured against current operations must show an increase in capacity, therefore a meaningful benefit, while maintaining current safety standards. Capacity measures should address all domains of the air traffic system from gate to gate.

6.1.2 Suggested Measures

Metric	Agent Domain
Aircraft throughput.	All
Aircraft time in sector.	ATSP
Aircraft (instantaneous count) in sector.	ATSP
Spacing on final.	ATSP and FC
Time between arrivals.	ATSP and FC
Average separation of aircraft for en route and terminal areas.	ATSP and FC
Frequency utilization: VHF or Datalink.	ATSP and FC

6.2 EFFICIENCY

6.2.1 Background

Through the implementation of DAG-TM concepts, the airline users hope to achieve an increase in flight efficiency in terms of time and fuel savings. In the literature, efficiency of air traffic has been measured objectively in terms of aircraft fuel burn, time, and distance (FAA & NASA, 2001; Wing, Adams, Duley, Legan, Baremore & Moses, in press). In terms of human performance, efficiency is measured by the extent to which an operator can handle a task (Galushka et al., 1995; Paas & van Merrienboer, 1993).

6.2.2 Suggested Measures

Metric	Agent Domain
Aircraft at optimal altitudes: number of aircraft on optimal altitudes, time and distance traveled on optimal altitudes.	All
Aircraft at optimal speeds: number of aircraft at optimal speeds, time and distance traveled at optimal speeds.	All
Distance between aircraft (CPA above separation standards).	ATSP and FC
Deviation from preferred route: time and distance.	FC and ATSP
Fuel burn.	FC
Delay: gate, take-off, arrival.	All
Time savings: time gain off average flight route.	All
Aircraft maneuvers: number and type.	ATSP and FC
Time deviation from required time of arrival (RTA).	FC
Self-spacing: deviation from target rate.	FC
Arrival rate.	ATSP and FC
Departure rate.	ATSP and FC

6.3 FLEXIBILITY

6.3.1 Background

A potential benefit of DAG-TM is system flexibility, allowing more degrees of freedom in flight planning for the FC and AOC. In today's environment, FC flexibility is limited by constraints of ATSP positive control, and inadequate information dissemination (weather, delays, etc). This concept is something that is frequently discussed in terms of free flight benefits, and is frequently a component of study; however, the method of flexibility measurement is an area that needs more study. Measures of flexibility include the ability to recover from errors or aircraft blunders, flight path deviations (Ozmore & Morrow, 1996; Wing et al., in press), and the transition of free flight between authorities (FAA & NASA, 2001). It is also dependent upon timeliness and extent of information transfer, available decision support tools, and trust in the roles of each separating authority. Standard measures have not been established, however, the Free Flight Phase 1 metrics team proposed to direct measures of the flexibility construct in terms of meeting the users' needs on an individual flight basis (FFP1, 1999). For example, an air carrier may prefer more delay for one of its flights in order to make up time for another, therefore the ability for the ATSP to accommodate these types of user requests would be one measure of flexibility. AOC input on intent and priorities would be extremely useful for study in this area.

6.3.2 Suggested Measures

Metric	Agent Domain	Measurement Method
Number of pilot requested deviations approved.	ATSP	Post processing of communications data.
Number of altitudes used.	All	Post processing of flight profiles as flown.
Number of altitudes not used.	All	Post processing of flight profiles as flown.
Number of aircraft at each altitude.	All	Post processing of flight profiles as flown.
Number of flights that used their optimal/desired altitudes (and duration at that altitude).	All	Post processing of flight profiles as filed and as flown.
Number of weather deviations.	All	Post processing of flight data.
Trajectory to optimize local flow constraints.	TFM, AOC	Fuel burn, time in flight, distance traveled.
Ability to “make up” time – reduction of individual aircraft delay.	All	Post processing of flight data.
Number of times arrival, departure, or en route slots are exchanged.	TFM, AOC, ATSP	Post processing of flight data and schedule.
Number of in-flight modifications to flight plan (requested by FC and approved by ATSP).	FC and ATSP	Post processing of flight and communications data.
Number of in-flight modifications to flight plan (initiated by ATSP and executed by ATSP).	ATSP and FC	Post processing of flight and communications data.

6.4 SAFETY

6.4.1 Background

System safety has been measured in terms of operational errors, aircraft proximity, conflict alerts, and errors (Allendoerfer & Galushka, 1999; Buckley et al., 1983). These measures have been widely used in experiments that involve ATSP and FC operations (e.g., FAA & NASA, 2001).

6.4.2 Suggested Measures

Metric	Agent Domain
Operational error • Loss of separation (distance and duration traveled with loss of separation), and • Airspace violations.	ATSP and FC
Conflict alerts • Number, • Duration, and • Type (system specific).	ATSP and FC
Error • Keystroke error, • Read back and hear back error, and • Missed handoff.	ATSP and FC
Closest point of approach • Safety issue below separation standard (up until standard it is a measure of efficiency).	ATSP and FC

6.5 WORKLOAD

6.5.1 Background

One definition of workload is the relationship between the imposed demands of the task and the availability of channel capacity or mental resources to deal with those commands. A substantial number of studies reviewed included measures of workload. Sanders and McCormick (1993) classified measures of mental workload into four categories: primary task (performance) measures, secondary task (performance) measures, physiological measures, and subjective measures. For the purpose of the guidelines, these measures were retained under a single construct.

Performance as a measure of workload assumes that as workload increases, performance decreases. According to O'Donnell and Eggemeier (1986), problems associated with this approach include: (a) underload may enhance performance; (b) overload may result in a floor effect; (c) confounding effects of information-processing strategy, training, or experience; and (d) measures are task specific and cannot be generalized to other tasks.

A widely used workload measure is the secondary task. A wide range of methods is presented in Gawron (2000). Advantages with this technique may be that it is a sensitive measure of operator capacity and stress. The possibility of implementation across tasks lends to ease of data comparison. Disadvantages may include task intrusion and operator's employing different strategies for the secondary task.

According to Casali and Wierwille (1983) subjective workload measures are often, inexpensive, unobtrusive, easily administered, and transferable. Wickens (1984) contends that these measures have high face validity. Limitations may include: (1) potential to confound mental and physical workload, (2) unconscious processing of information that the operator cannot rate subjectively, and (3) dependence on short-term memory (O'Donnell & Eggemeier, 1986). In addition, this method relies on well-defined questions to circumvent raters from interpreting questions and rating scales differently.

6.5.2 Suggested Measures

6.5.2.1 Workload: Primary Task Measure

Construct:	Workload: performance measure – primary task.
Metric:	Control Movements/Unit of Time.
Description:	The number of control inputs made summed over each control used by one operator divided by the unit of time over which the measurements were made.
Requirements:	System recordings of control movements and frequency of feature usage.
Strengths:	Highly sensitive workload performance measure.
Limitations:	Control movement must be defined.
Implementation:	FC: average count per second of inputs of flight controls. Other flight deck manipulations (glass cockpit features, etc.). ATSP: frequency and duration of time engaged in communications, keyboard usage, data block management, etc.
Acquisition Information:	N/A.
Estimated Cost:	N/A.
References:	Wierwille and Conner, 1983. Porterfield, 1997.

6.5.2.2 Workload: Secondary task measure

Construct:	Workload: performance measure – secondary task.
Metric:	Secondary Task.
Description:	Requires an operator to perform the primary task within that task's specified requirements and to use any spare attention or capacity to perform a secondary task. The decrement of performance of the secondary task is operationally defined as a measure of workload.
Requirements:	Secondary task vary widely. Task may or may not be related to primary tasks. Secondary tasks used in past research have included math problems, card sorting, or symbol identification, etc. Task should not interfere with the primary task, should be easy to learn, can be recorded or scored continuously.
Strengths:	Common metric for various tasks.
Limitations:	May interfere with primary task. Subjects may include the secondary task in any subjective workload rating (if both measures are used concurrently).
Implementation:	Task should be concurrent with primary task, and attempt to tap the same mental resources as the primary task.
Acquisition Information:	N/A.
Estimated Cost:	N/A.
References:	Gawron, 2000.

6.5.2.3 Workload: Physiological Measure

Construct:	Workload: physiological measure.
Metric:	Pupillary response.
Description:	Measurement of the pupil dilation during task performance.
Requirements:	Ocular measurement equipment.
Strengths:	Continuous data collection capability.
Limitations:	Cumbersome equipment must be attached to participant.
Implementation:	Magnetic Head Tracker. Equipped laboratories located in NASA LaRC and FAA Technical Center RDHFL 609-485-6152.
Acquisition Information:	UserWorks, Inc. 1738 Elton Road, Suite 138 Silver Spring, MD 20903 Fax: 301-431-4834 e-mail: info@userworks.com
Estimated Cost:	
References:	Willems, Allen, and Stein, 1999.

6.5.2.4 Workload: Subjective measures

Construct:	Workload: subjective measure.
Metric:	Air Traffic Workload Input Technique (ATWIT) or Pilot Objective/Subjective Workload Assessment Technique (POSWAT).
Description:	Subjective workload measured at standard intervals during the simulation. Array presents up to 10 input buttons. Each key corresponds to a workload rating. At predetermined intervals, low-level beep is emitted and lights of a numbered Workload Assessment Keyboard (WAK) are illuminated for a specified time interval (i.e., 20 seconds), after which, participants are instructed to enter a subjective workload rating. If the participant is unable to enter a rating by the end of the allotted time, an automatic rating of 99 is recorded. POSWAT version also measures entry reaction time.
Requirements:	Workload Assessment Keyboard (WAK) (see Figure 1) or other functional equivalent * (see Figure 2) input device.
Strengths:	High reliability, real-time ratings, variable rating interval
Limitations:	Minimal disruption of task.
Implementation:	Real time, portable, instantaneous data extraction.
Acquisition Information:	A & J Industries, (405) 794-6667. POC: Alie Burgin.
Estimated Cost:	$3,000.00 (four WAK units and multiplexer). Programming is required.
References:	FAA & NASA 2001; Porterfield, 1997; Sollenberger, Stein & Gromelski, 1997; Stein, 1984.

Figure 1. Workload Assessment Keypad (WAK).

Construct:	Workload: subjective measure.
Metric:	Subjective workload rating.
Description:	*Functional equivalent of the WAK (see Figure 2).
Requirements:	If possible, an input keyboard can be incorporated directly on the participants interface. Workload data can be recorded and time stamped with system data recording.
Strengths:	High reliability, real-time ratings, variable rating interval.
Limitations:	Minimal disruption of task.
Implementation:	Programming (i.e. power point, visual basic) into system interface required.
Acquisition Information:	Developer: Mark Peters: mpeters@seagull.com NASA Langley Research Center Contact: David Wing: Tel: (757) 864-3006.
Estimated Cost:	N/A.
References:	Wing et al., in press.

Figure 2. Workload Control Panel.

Construct	Workload: subjective measure
Metric:	NASA TLX
Description:	A multi-dimensional rating procedure that provides an overall workload score based on a weighted average of ratings on six subscales: mental demands, physical demands, temporal demands, own performance, effort, and frustration.
Requirements:	Paper and pencil package (currently being updated) or a PC compatible program available. Participants rate six subscales (see Figure 3), and then perform 15 pairwise comparisons of six workload scales. Separate weights should be derived for diverse tasks.
Strengths:	High reliability. Used extensively in aviation research. May be used in operational environments.
Limitations:	Requires weighting scales.
Implementation:	Paper or PC based rating scale.
Acquisition Information:	Currently maintained by: Human Systems Information Analysis Center. http://iac.dtic.mil/hsiac/products/tlx/tlx.html.
Estimated Cost:	N/A.
References:	Hart and Staveland, 1987.

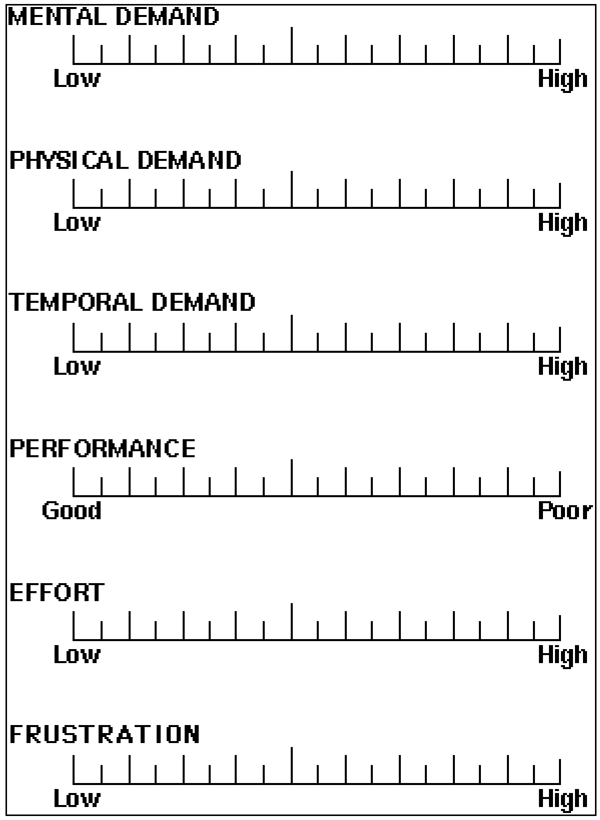

Figure 3. NASA TLX Rating Sheet

Construct:	Workload: subjective measure.
Metric:	Subjective Rating Questionnaire (i.e., Cooper Harper Rating Scale).
Description:	Ordinal scale questionnaire.
Requirements:	Subjects complete workload ratings on a post scenario or simulation questionnaire.
Strengths:	Easy to administer and widely accepted.
Limitations:	Potential confound of mental and physical workload.
Implementation:	Post scenario and/or sim. Overall rating. Non-parametric statistics.
Acquisition Information:	N/A.
Estimated Cost:	N/A.
References:	Allendoerfer and Galushka, 1999.

6.6 DECISION MAKING

6.6.1 Background

New roles and responsibilities that result from DAG-TM concepts will require different decision-making styles, decision support tools and procedures. Since these changes will alter the human tasks, the human/machine interface, and the allocation of functions between these agents, it is imperative that we examine the roles and responsibilities of air/ground traffic agents and assess the impact of these changes on the air transport system.

Several studies have investigated the decision making process for the ATSP (Endsley, Mogford, Allendoerfer, Snyder, and Stein, 1997; Endsley 1997; Fleming, Lane, and Corker, 2000) and the FC (Cashion & Lozito, 1999; Mackintosh et al. 1998) in a shared separation environment. Decision making assessments include the Controller Decision Evaluation (Borg, 1978; Kinney, 1977), and the Critical Incident Technique (Flanagan, 1954). Decision making processes are not well known for the AOC. The FAA and NASA study (2001), examined decision making shared separation environment, however, measures across all air traffic agents need further definition.

6.6.2 Suggested Measures

Metric	Agent Domain	Measurement Method
Time to make a decision time between conflict alert and plan development	All	Real-time event data collection tool.
Time to implement a decision time between plan development and plan implementation	All	Real-time event data collection tool.
Time for information exchange time between plan implementation to intent or request broadcast response time of other agent to that intent or request	All	Observer event log, real-time event data collection tool.
Decision impact on workload	All	Subjective/ questionnaire.
Decision impact on efficiency	All	Observer log, communications analysis, flight data, or questionnaires.
Decision impact on flexibility	All	Observer log, communications analysis, flight data, questionnaire data.
Decision impact on safety	All	Observer log, communications analysis, flight data, questionnaire data.
Decision impact on complexity	All	Observer log, communications analysis, flight data, questionnaire data.

6.7 SITUATION AWARENESS

6.7.1 Background

For the DAG-TM concept, situation awareness (SA) is a critical construct since it is an integral component of decision-making. SA has been defined by Endsley (1988) as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status on the near future". There is a great level of effort to identify decision making authority roles and responsibilities to agents across the ATM triad, information sharing requirements, and automation requirements to ensure safe transfer of flight authority. SA must be maintained between the agent and the system, between the agent and the system's relationship to the environment (does the system perform differently under certain conditions), and between agents. Gawron (2000) lists three types of SA measures: performance, subjective ratings, and simulation.

SA is measured through performance, query methods, subjective ratings, and simulation. Probably the most well known method of measuring SA is the Situational Awareness Global Assessment Technique (SAGAT) (Endsley, 1988) that uses real-time, human-in-the-loop simulations during which the scenario is stopped randomly and the operators are queried. Sarter and Woods (1991) contend that this method measures recall, not SA. The following SA measures each have some strengths and weaknesses, therefore a researcher should select a measure based on convenience, ease of use, and relevance to the study. If freezing a scenario is an option then SAGAT would be a potential SA measurement choice. If freezing is not an option, then supplemental questionnaires (i.e., SART) would be a good candidate. Irrespective of a method, objective performance data that reflects SA (or lack thereof) should collected.

6.7.2 Suggested Measures

6.7.2.1 Situation Awareness: Performance Measures

Construct:	SA: performance measure.
Metric:	Situational Awareness Global Assessment Technique (SAGAT).
Description:	Designed for real-time, human in the loop simulations. Simulation is stopped at random intervals and the participants are queried about characteristics of the scenario or system at that point in time. Answers are compared to the data collected in the system at the same point in time.
Requirements:	Queries must be predetermined. Questions may be programmed to appear on the user's screen and entered dynamically, or verbally queried by a researcher.
Strengths:	Objective measure of percent correct or errors.
Limitations:	Subject to memory decay and inaccurate beliefs. Requires the capability to pause the simulation.
Implementation:	Real time. System must allow repeated mid-simulation pauses.
Acquisition Information:	N/A.
Estimated Cost:	N/A.
References:	Endsley, 1988.

In addition, objective performance measures of SA may include:

Metric	Agent Domain
Missed hand-offs.	ATSP
DSR J-ring (halo) initiation (* HOST records this as a toggle action and does not indicate on/off. Analysis may be difficult if using SAR data only).	ATSP
Time to accept hand-offs or acknowledge point outs.	ATSP
Control panel feature usage to gain information.	All agents
Detection of a traffic conflict before receiving an automated alert.	ATSP & FC

6.7.2.2 Situation Awareness: Subjective measures

Construct:	SA: subjective measures.
Metric:	The Situation Awareness Subjective Workload Dominance Technique (SWORD).
Description:	Evaluation of alternate measure of SA on task using pairwise comparisons.
Requirements:	Develop a rating scale listing all possible pairwise comparisons for each performance task, the subject completes an awareness judgment matrix comparing each task pair, and ratings are calculated using geometric means.
Strengths:	Unknown.
Limitations:	Subjective reliability, subject memory decay.
Implementation:	Rating scale development, task identification.
Acquisition Information:	N/A.
Estimated Cost:	N/A.
References:	Fracker & Davis 1991.

Construct:	SA: subjective measure.
Metric:	Situational Awareness Rating Technique (SART).
Description:	A questionnaire that measures across three domains: demands on attentional resources, supply of attentional resources, and understanding of the situation (see Figure 4).
Requirements:	Administer questionnaire measuring operator knowledge.
Strengths:	Easy to administer. Developers evaluated SART in terms of aircraft attitude recovery tasks and learning comprehension. Developers recommend this method for comparative system design evaluation.
Limitations:	Subjective reliability.
Implementation:	Subjective ordinal scale ratings. Use non-parametric statistics for analysis.
Acquisition Information:	N/A.
Estimated Cost:	N/A.
References:	Taylor, 1990.

		Low						High
		1	2	3	4	5	6	7
Demand	Instability of Situation							
	Variability of Situation							
	Complexity of Situation							
Supply	Arousal							
	Spare Mental Capacity							
	Concentration							
	Division of Attention							
Under-standing	Information Quantity							
	Information Quality							
	Familiarity							

Figure 4. SART Rating Scale

Table 3. Definitions of SART Rating Scales (Taylor & Selcon, 1991).

Instability:	Likelihood of situation changing suddenly.
Complexity:	Degree of complication of situation.
Variability:	Number of variables changing in situation.
Arousal:	Degree of readiness for activity.
Concentration:	Degree of readiness for activity.
Division:	Amount of attention in situation.
Space Capacity:	Amount of attentional left to spare for new variables.
Information Quantity	Amount of information received and understood.
Information Quality	Degree of goodness of information gained.

6.7.2.3 Situation Awareness: Simulation measures

Construct:	SA: simulation measures.
Metric:	Computational Situational Awareness model (CSA).
Description:	Three computational components including: situational element (SE), context-sensitive nodes, and a regulatory mechanism that assesses the situational elements for all nodes.
Requirements:	Embedded in MIDAS model.
Strengths:	The systems model includes the cockpit, or workstation, the environment, and the human figure model. CSA data had a high correlation with SART data.
Limitations:	Unknown.
Implementation:	Each SE has a mathematical weight based upon its importance in the situation and a mathematical value based upon one of four quantifiable levels of awareness (detection, recognition, identification, and comprehension). Situation-sensitive nodes are semantically related collections of SE's. The nodes are defined by the context of a given task and are weighted by the overall importance of the node in determining the level of SA. If the situation changes, then the weights on the nodes, or the nodes themselves, may change to reflect accurately the level of SA. SA is the weighted average of knowledge that the pilot has in each node, and thus is a measure of the pilot's perceived SA. The CSA model then subtracts an error component, based on misidentified SE's or unknown elements in the environment.
Acquisition Information:	Point of Contact: R. Shively (650) 604-6249. jshively@mail.arc.nasa.gov.
Estimated Cost:	N/A
References:	Shively, Brickner, & Silbiger, 1997.

6.8 USABILITY

6.8.1 Background

Usability evaluation of a system is generally subjective in nature and may be assessed comprehensively by utilizing rating scale and interview techniques. In addition, objective measures such as accuracy and reaction time to complete a task may be employed for system design comparisons.

Measures of usability are usually system specific, but are generally measured subjectively in terms of ease of use, display characteristics and effectiveness, and workstation layout (Allendoerfer & Galushka 1999; American National Standards Institute, 1993). The AHP (Saaty, 1980) may be a suitable method for task performance usability.

6.8.2 Suggested Measures

Construct:	Usability.
Metric:	Rating scale questionnaire or structured interview.
Description:	Questions should elicit user opinion on: ease of use, accessibility of all necessary controls, ease of reading, ease of understanding, function, limitations, effectiveness, and overall information presentation and user interaction.
Requirements:	N/A.
Strengths:	May be tailored to encompass system specific functions.
Limitations:	User participants for DAG-TM must be subject matter experts. Questions and ratings must cover all critical aspects of the system.
Implementation:	Subjective ordinal scales. Non-parametric analysis required.
Acquisition Information:	N/A.
Estimated Cost:	N/A.
References:	Allendoerfer & Galushka, 1999.

6.9 COMPLEXITY

6.9.1 Background

The calculation and prediction of dynamic density (DD) has been identified as a key need for assessing workload of future traffic (RTCA Task Force 3 Free Flight Implementation Report, 1995). The monitor alert parameter is the current measure of airspace complexity, however, this is based solely on aircraft count and is not considered useful for projective ATC planning. A collaborative effort by the FAA, NASA, METRON, and other agencies to develop an improved measure is underway. The FAA has imposed a requirement to deploy an improved metric (either one of the proposed metrics or a combination of metrics from various agencies) by early 2002. This metric will be validated using the Enhanced Traffic Management System (ETMS) data, which is currently a NAS wide application. If the effort is continued, the logical next step is to validate a complexity metric using Center TRACON Automation System (CTAS) data, which has better prediction capability.

Mogford, Guttman, Morrow, and Kopardekar (1995) did an extensive review all factors that contribute to the complexity. Their review provided a foundation for developing dynamic density variables that mathematically represent the effect of these factors. In a review of proposed variables, Kopardekar (2000) presented a review of complexity metrics developed by in various studies. He identified common factors that contribute to air traffic complexity citing aircraft count, sector geometry, traffic flows, separation standard, aircraft performance characteristics, weather as the most common factors that contribute to the air traffic complexity or difficulty. This review includes a comprehensive list of variables developed by several researchers (see Table 4).

6.9.2 Suggested Measures

Once a common metric has been established, scenarios across various types of studies (fast time, real-time human in the loop) may be evaluated and compared based on a standardized complexity rating. A standardized metric will also provide researchers with a benchmark to evaluate the impact of variables such as new tools and procedures on airspace complexity. These proposed DD metrics are based largely on the air traffic control perspective; therefore, it will be necessary to modify this metric slightly for flight deck centered simulations. The DD measures are currently being validated by the FAA WJHTC. The validation activity will examine which measures best capture complexity. Once the validation activity is completed, a smaller set of these measures will be recommended.

Table 4. Dynamic Density Variables

ID	Variable Name	Description	Purpose
1	*Kopardekar, P., and DiMeo, K, 1997, Dynamic Density Variables, PowerPoint Briefing.*		
1.1	Aircraft density (AD)	Number of aircraft divided by the airspace volume. See formula. [1] $$AD = \frac{Number\ of\ Aircraft}{Airspace\ Volume}$$	Capture the monitoring taskload.
1.2	Convergence recognition index (CRI)	Based on how close the angle of convergence is to the shallow convergence of 30 ^{0}. See formula. [1] $$CRI = \left(7 - \frac{Adjusted\ Convergence\ Angle}{30}\right)^2$$	Capture the degree of difficulty in recognizing shallow convergence angles.
1.3	Separation criticality index (SCI)	Based on how close the separation between the two aircraft will be in relation to the separation minima. See Formula. [1] $$SCI = (3 - SI)^2$$ $$Separation\ Index\ Longitudinal\ (SILO) = \frac{\Delta Y}{Logitudinal\ Minima}$$ $$Separation\ Index\ Lateral\ (SILA) = \frac{\Delta X}{Lateral\ Minima}$$ $$Separation\ Index\ Horizontal\ (SIH) = \frac{\sqrt{\Delta X^2 + \Delta Y^2}}{Lateral\ Minima}$$ *En Route and Terminal* $$Separation\ Index\ (SI) = \frac{SIV + SIH}{2}$$	Capture the monitoring and decision making taskload associated with examining separation criticality. Include if *SIH* < *4*, *SIV* < *2, and SI* < *3.*

Table 4. Dynamic Density Variables (Continued)

ID	Variable Name	Description	Purpose
1.4	Degrees of freedom index (DOFI)	Based on how many options the controller has to move the aircraft that are in a potential conflict situation. See Formula. [1] $DOFI = (12 - Available\ DOF\ per\ pair)^2$ *DOF = Degrees of Freedom*	Capture the degree of difficulty in developing resolution strategies for conflict pair. Fewer the options, the higher the difficulty.
1.5	Coordination taskload index (CTI)	Based on how close the aircraft is to the sector boundary that is not handed off to the next sector. See Formula. [1] $CTI = \left[\frac{1}{Time\ to\ reach\ CCP + (Time\ to\ reach\ SB - Time\ to\ reach\ CCP)^2} \right]$ $CTI = [10 - Time\ to\ reach\ Sector\ Boundary]^2$ CCP = Coordination Completion Point, a point at which facility procedures require that coordination must be completed (e.g., 2.5 miles from sector boundary).	Capture the taskload associated with the coordination. The larger the distance from the sector boundary, the lower the urgency for hand-off and lower the taskload. Smaller the distance from the sector boundary, higher the urgency for hand-off and higher the taskload. All aircraft, which are 10 minutes or less from CCP are included.
2	*Breitler, A., L, Lesko, M. J., Kirk, K. M., 1996, Effects of Sector Complexity and Controller Experience on Probability of Operational Errors in Air Route Traffic, CAN Corporation, Task 9, FAA Contract DTFA01-95-C-00002.*		
2.1	Operationally acceptable level of traffic (OALT)	Maximum number of aircraft in a sector that a controller can manage effectively during a defined time period. [2]	Rule of thumb based on number of aircraft. This is the ETMS alert parameter. A static variable. Correlation with operational errors (r = -0.06, N.S.)
2.2	Maximum instantaneous aircraft (MIAC)	Maximum number of aircraft that a controller should ever have to manage in the sector during a defined time interval (60 minutes). [2]	Similar to OALT but based on time period. Static parameter. Correlation with operational errors (r = - 0.104, N.S.)
2.3	Ease of vectoring	Subjecting rating by each center (from 1 to 5). [2]	ARTCC wide vectoring ease. Static parameter. Correlation with operational errors (r = 0.184, S.)

Table 4. Dynamic Density Variables (Continued)

ID	Variable Name	Description	Purpose
2.4	Ease of transitioning	Subjective rating by each center (from 1 to 5). [2]	ARTCC wide transitioning ease. Static parameter. Correlation with operational errors (r = 0.231, S.)
2.5	Number of sides	Number of sides for each sector. [2]	Sector geometry related difficulty. Static parameter. Correlation with operational errors (r = 0.094, N.S.)
2.6	Number of main jetways	Number of main jet routes through each sector. [2]	Capture sector flow. Static parameter. Correlation with operational errors (r = -0.061, N.S.)
2.7	Number of fixes/airports	Number of fixes and controlled airports in each sector. [2]	Capture sector operation. Static parameter Correlation with operational errors (r = 0.219, S.)
2.8	Letters of agreement (LOAs)	Number of letters of agreement for each sector. [2]	Capture sector operation. Static parameter. Correlation with operational errors (r = -0.162, S.)
2.9	Traffic count (TC)	Number of aircraft in the sector within 15-minute time period. [2]	Capture traffic load. Based on ETMS data. Dynamic parameter.
2.10	Traffic level (TL)	Number of aircraft in sector in the previous hour. [2]	Capture traffic load after effects from previous hour. Based on ETMS data. Dynamic parameter.
2.11	Entries	Number of new aircraft in the sector within 15-minute time periods. [2]	Capture traffic load due to new arrivals. Based on ETMS data. Dynamic parameter.
2.12	Average ground speed (GS)	The average ground speed of all aircraft in the sector within 15-minute time period. [2]	Capture taskload due to Ground speeds. Based on ETMS data. Dynamic parameter.

Table 4. Dynamic Density Variables (Continued)

ID	Variable Name	Description	Purpose
2.13	Standard deviation of average ground speed	The standard deviation of ground speed within 15-minute time period. [2]	Capture taskload due to ground speed differential. Based on ETMS data. Dynamic parameter
2.14	Average altitude (ALT)	The average altitude of aircraft in the sector within 15-minute time interval. [2]	Capture taskload due to altitudes used. Based on ETMS data. Dynamic parameter.
2.15	Standard deviation of average altitude	Measurement of the distribution of altitudes of aircraft within the sector in 15-minute intervals. [2]	Capture taskload due to altitude differential. Based on ETMS data. Dynamic parameter.
2.16	Course bins	A count of the number of different increments of 20 degrees in bearing which are occupied by at least one aircraft in the sector for each 15-minute period. This is a measure of the complexity of traffic with regard to course within the sector. [2]	Capture the bearing difference in the sector. Based on ETMS data. Dynamic parameter. (I am not sure what this measure means)
2.17	Transitions	Number of aircraft changing altitude in a 15-minute period. [2]	Capture taskload due to transitions. Based on ETMS data. Dynamic parameter.
	Complexity	*= TC + TC + AVG_GS + STD_GS + STD_ALT + TR + STD_BETA* [2]	Capture complexity at sector level.
3	*Chatterji, G. B., Shridhar, B, 1997, Measures of Airspace Complexity, Preliminary Draft and Unpublished Work, NASA Ames Research Center.*		
3.1	Maximum terrain elevation	Fixed number [3]	Static
3.2	Usual cloud ceiling	Fixed number [3]	Static
3.3	Volume of airspace available	Fixed number [3]	Static
3.4	Number of merging points	Fixed number [3]	Static
3.5	Number of neighboring sectors that hand-off traffic	Fixed number [3]	Static

Table 4. Dynamic Density Variables (Continued)

ID	Variable Name	Description	Purpose
3.6	Number of neighboring sectors that accept traffic hand-offs	Fixed number [3]	Static
3.7	Sector operating procedures	Fixed number [3]	Static
3.8	Navigational aids available	Fixed number [3]	Static
3.9	Surveillance equipment available	Fixed number [3]	Static
3.10	Aircraft mix	See formula [3]	Dynamic
3.11	Traffic density	See formula [3]	Dynamic
3.12	Available airspace	See formula [3]	Dynamic
3.13	Operations near reserved airspace	See formula [3]	Dynamic
3.14	Operations near sector boundaries	See formula [3]	Dynamic
3.15	Flight mode	See formula [3]	Dynamic
3.16	Onboard equipment	See formula [3]	Dynamic
3.17	Horizontal proximity	See formula [3]	Dynamic
3.18	Vertical proximity	See formula [3]	Dynamic
3.19	Time-to-go	See formula [3]	Dynamic
3.20	Groundspeed variability	See formula [3]	Dynamic
3.21	Conflict resolution	See formula [3]	Dynamic
3.22	Preferred path	See formula [3]	Dynamic
3.23	Shape of traffic geometry	See formula [3]	Dynamic

Table 4. Dynamic Density Variables (Continued)

ID	Variable Name	Description	Purpose
4	*Position Classification for Air Traffic Control, Series ATC 2152, 1997, Draft*		
4.1	Sustained traffic index (Dt)	*Dt* = 1 + (*Cav2*/*Cav1*) *Cav1* is the average unweighted hourly count for busiest 1830 hours. [4] *Cav2* is the average unweighted count for the second busiest 1830 hours. [4]	Used for classifying ARTCC or TRACON business/complexity. Post processing is necessary.
4.2	Proposed ATC classification system	Based on weights for each type of operation (e.g., radar, VFR, jet traffic, prop/turbo-prop combination). [4]	Used for classifying ARTCC or TRACON business/complexity. Do not know if this was accepted. Need post processing
5	*McNally, B.D., Laudeman, I.V., Mayhugh, B., 1997, Field Test Evaluation Plan for a Conflict Prediction and Conflict Resolution System, NASA Ames Research Center.* *Shridhar, B., Seth, K. S., Grabbe, S., 1998, Airspace Complexity and its Application in Air Traffic Management, 2nd USA/Europe Air Traffic Management R&D Seminar.*		
5.1	Traffic density (TD)	Number of aircraft in the sector. [5]	Weight = 1.0
5.2	Heading change (HC)	The number of aircraft that made a heading change of greater than 15 degrees during the analysis time interval. [5]	Weight = 2.4
5.3	Speed change (SC)	The number of aircraft that had a computed airspeed change of greater than 10 knots or 0.02 mach during the analysis time interval. [5]	Weight = 2.45
5.4	Altitude change (AC)	The number of aircraft that made an altitude change of greater than 750 feet during the analysis time interval. [5]	Weight = 2.94
5.5	Speed differential (SD)	The number of aircraft that had a speed difference of greater than 150 knots from the average speed in the sector during the analysis time interval. [5]	Weight = 3.72
5.6	Minimum distance 0-5 nm (MD 0-5)	The number of aircraft that had a Euclidean distance of 0-5 nm to the closest other aircraft at the end of the analysis time interval, excluding the conflicting aircraft. [5]	Weight = 2.45
5.7	Minimum distance 5-10 nm (MD 5-10)	The number of aircraft that had a Euclidean distance of 5-10 nm to the closest other aircraft at the end of the analysis time interval, excluding the conflicting aircraft. [5]	Weight = 1.83

Table 4. Dynamic Density Variables (Continued)

ID	Variable Name	Description	Purpose
5.8	Conflict predicted 0-25 nm (CP 0-25)	The number of aircraft predicted to be in conflict with another aircraft whose lateral distance at the of the analysis interval is 0-25 miles. [5]	Weight = 4.00
5.9	Conflict predicted 25-50 nm (CP 25-40)	The number of aircraft predicted to be in conflict with another aircraft whose lateral distance at the of the analysis interval is 25-40 miles. [5]	Weight = 3.00
5.10	Conflict predicted 40-70 nm (CP 40-70)	The number of aircraft predicted to be in conflict with another aircraft whose lateral distance at the end of the analysis interval is 40-70 miles. [5]	Weight = 2.11
	Complexity =	Weighted sum of all the above variables.	Composite dynamic density metric.
6	*Knecht, W., Smith, K., and Hancock, P.A., 1996, A dynamic conflict probe and index of collision risk, Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting.*		
6.1	Index of collision risk	*Dt* = $\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\frac{1}{\left[dij(t)/c\right]^{a}}$ *N* = Number of aircraft *dij* = distance between two aircraft not separated by altitude *c* = normalization constant, equal to 5 nm, the minimum allowable lateral separation *a* = weighting factor set to 3 [6]	Based on separation as the most important factor in collision risk estimation.
7	*Wyndemere, 1996, An Evaluation of Air Traffic Control Complexity, Final Report, Contract Number NAS2-14284 (NASA contract).*		
7.1	Aircraft count (ACT)	Number of aircraft in a sector. [7]	
7.2	Angle of convergence in conflict situation (ANG)	Measurement of the severity of each conflict situation based on the conflict geometry. [7]	Conflicts with small convergence angles are difficult to handle.
7.3	Number of aircraft climbing or descending (CoD)	Count of the number of aircraft that are in climb or descent at an instant in time. [7]	

Table 4. Dynamic Density Variables (Continued)

ID	Variable Name	Description	Purpose
7.4	Closest point of approach (CPA)	Weighting of the number of aircraft that are within a threshold separation of each other at any instant in time. [7]	CPA thresholds are between 8 and 13 miles. One unit for aircraft that are below 8 CPA and 0.5 unit for aircraft that are between 8 and 13 CPA.
7.5	Aircraft density (DNS)	Aircraft count divided by the usable amount of sector airspace. [7]	
7.6	Level of knowledge of intent of aircraft (INT)	Level of information about the intent. Three categories: current operations, half free flight, and full free flight. [7]	Zero complexity for current operations, 0.5 for half free flight, and 1 for full free flight.
7.7	Neighbors (NBR)	Aircraft that are predicted to be within a threshold separation. [7]	One unit for each aircraft that is within 10 lateral miles and 2000 vertical feet.
7.8	Proximity of aircraft to sector boundary (PRX)	Count of the aircraft that are within a threshold distance of a sector boundary. [7]	Greater coordination and monitoring is required when aircraft are closer to the sector boundary.
7.9	Proximity of potential conflicts to sector boundary (PRX-C)	Count of the predicted conflicts that will occur within a threshold distance of a sector boundary. [7]	Controller will have less time to resolve a conflict situation that is near a sector boundary. One unit for each conflict that is within 10 miles of sector boundary and 0.5 unit for each conflict that is within 20 miles of the sector boundary.
7.10	Airspace structure (STR)	Measure of conformance of the traffic flow through a sector to the geometry of the sector. Calculations using major axis, aspect ratio, and difference in heading and the major axis. The squared deviation from the major axis of the sector is weighted by the aspect ratio and then summed over all aircraft. [7]	Complexity may increase if majority of aircraft fly against the grain.
7.11	Variance in aircraft speed (VAS)	Measure of the variability of ground speed of all aircraft in the sector. (e.g., standard deviation) [7]	

Table 4. Dynamic Density Variables (Continued)

ID	Variable Name	Description	Purpose
7.12	Variance in directions of flight (VDF)	Measure of the variability of heading of all aircraft in the sector. [7] $\frac{1}{(n)(n+1)} \sum_{i} \sum_{j,j\neq i} \left((hdg_i - hdg_j)^2 \right)$	A higher heading variability of the traffic situation provides less organization of the traffic flow. Controller can group individual aircraft together with lower heading variability.
7.13	Airspace structure (STR)	Description of sector size and structure. [7]	Capture sector geometry related difficulty (narrow, long, etc.)
7.14	Crossing altitude Profiles (CAP)	Count of number of aircraft pairs in which one aircraft will be climbing and one aircraft will be descending through the same altitude. [7]	
7.15	Altitude variation (VAA)	Measure of the variability of altitude of all the aircraft in the sector. [7]	No evidence from SMEs about its validity.

6.10 SIMULATION FIDELITY

6.10.1 Background

There is currently no standard metric for simulation fidelity. However, like complexity, it is an integral variable in research assessments. Limitations in simulation fidelity may have a severe impact on study results. Care should be taken to ensure that whatever limitations may exist in a simulation (inability to emulate winds, inaccurate aircraft performance, etc.) are factored into the results of the study.

6.10.2 Suggested Measures

Subtask 5 Simulation Requirements guidelines document (Kopardekar, et al., 2001) outlines three approaches for simulation fidelity assessment: fidelity based on general classification, fidelity based on adequacy of a simulator, and fidelity based on quantitative approach. Please refer to this document for specific simulation fidelity assessment methods.

Allendoerfer and Galushka (1999), provide a list of recommended simulation fidelity metrics to evaluate whether all other data elements in a simulation were collected under realistic conditions. These metrics also provide a basis of comparison across simulations. Table 5 provides a list of these metrics with their definition and source.

Table 5. Simulation Fidelity Metrics

Metric	Definition	Source
Traffic Scenario Characteristics	Length of scenario, Average number of aircraft entering the scenario/minute, Total number of arrivals, Total number of departures, Total number of overflights, Total number of propeller aircraft, Total number of jet aircraft, and Total number of scripted pilot deviations and requests.	Scenario system development recording.
Other Simulation Characteristics	Standard operating procedures and letters of agreement used, and Flight strip timing.	Overall.
Realism Rating	Perceived realism and fidelity of the simulation as rated by the participant.	Questionnaire.
Impact of Technical Problems	Perceived impact of technical problems on the participants' ability to complete simulation tasks.	Questionnaire.
Impact of Pseudopilots Rating	Perceived impact of the pseudopilots on the participants' ability to complete simulation tasks.	Questionnaire.
Scenario Difficulty Rating	Perceived difficulty of traffic scenario.	Questionnaire.

Additional metrics that do not fit under the proceeding sections include trust in automation and trust in other agents. Further metric development is needed in this area; however, there are examples of trust data collection. In the AUTRII experiment (Wing et al. in press), pilot participants could select an automated conflict resolution flight path or choose to fly manually. If a participant consistently chooses to disregard automation help, this would be an indication of lack of trust in automation. In the AGIE experiment (FAA & NASA, draft report), controller participants indicated a lack of trust in pilot conflict resolution abilities when they canceled free flight operations in a shared separation condition. In the debriefing sessions, controllers indicated that they were not confident that the pilot's conflict resolution would maintain standard separation (this included conflict aircraft flight paths scripted to maintained separation).

7. EXPERIMENTAL DESIGN

The following section outlines general considerations for experimental design. However, for detailed information concerning experimental design a number of texts offer more in depth information (Keppel, 1982; Meister, 1986; and Montgomery, 1976); in addition Subtask 3: Human Factors Evaluations of NASA DAG-TM Concept Elements and Decision Support Tools (Kopardekar & Sacco, 2001) discusses this subject in greater detail.

7.1 DEFINE THE QUESTION

The first step in experiment planning involves clearly defining the questions to be addressed. Clear objectives will provide direction and focus, and will keep the scope of the study within realistic boundaries. Exploratory studies may be less defined, with the expectation that more questions will be the result of study efforts. However, when clear objectives call for more concrete results, a study should define the independent variables (what is measured), develop conditions in which to compare the variables, and select dependent variables (measures used for comparison). Among other specified conditions, the basic experimental design of DAG-TM studies should include a baseline condition of current ATM to ensure that automation, procedures, and concepts demonstrate an increase of efficiency, capacity, and other important considerations, while maintaining at least the same level of safety.

7.2 IDENTIFY LIMITATIONS

Limitations are the independent variables that restrict the ability to generalize the results of the study. Limitations in DAG-TM studies may include simulation fidelity, such as the lack of weather or winds. The participant population will also affect the ability to generalize results. Care should be taken to ensure that participants accurately represent the users of the system within the appropriate domain. A common problem in ATM studies is the extremely small sample size due to time, cost, and resource restraints.

7.3 SELECT MEASURES

Once objectives and conditions for a study have been determined, a set of constructs and metrics should be selected to adequately gather relevant information pertaining to the hypotheses. The methodology for these metrics should be outlined, including what is to be collected, the method of collection and any necessary calculations for data output files, where and how (what format) the data will be stored, and how the data will be analyzed. It is also important to incorporate data collection during "dry runs" to ensure that all components are recorded as intended and to remedy any problems before the study actually begins. This will also ensure that recording parameters are set to detect a meaningful difference in the measured variable. These recorded parameters should be organized in predefined file formats, with discreet naming conventions that may include the name of the study, date, run number, etc. Attention to the format of the data collection will help simplify the data analysis phase of the study.

7.4 GENERAL SCENARIO CONSIDERATIONS

Given the defined experimental conditions, other considerations will include: training, learning, the duration of the scenario, the level of complexity, and the environmental conditions. Adequate training will provide participants with laboratory familiarization, as well as the opportunity to get accustomed to any new procedures, concepts, or decision support tools. Scenario conditions should be similar enough to be comparable, but should be different enough to minimize a learning effect. In general, the length of the scenario should try to strike a balance between the need to gather required data and participant fatigue. The level of difficulty should be at least high enough to keep the participants engaged, and may be increased to see if performance degrades in the various conditions. Laboratory simulations should strive to emulate real world environmental conditions. Lighting, temperature, and noise should not distract participants from their given tasks.

8. FUTURE EFFORTS

Future efforts will include task specific recommendations. For each task, a battery of metrics will be presented across constructs necessary to assess the benefits and feasibility of the task. In Table, the example task is separation assurance. In order to adequately assess the feasibility of maintaining separation assurance in a distributed management scenario, measures should be made across several constructs. In many cases, more than one metric may be necessary or useful for a given construct. For example, metrics for safety may include the number of operational errors, closest point of approach, and communication delay. The suggested metrics will be hyperlinked to the metric guidelines exemplified in Table 6.

Table 6. Example of Metric Mapping

CE-5	Task: Separation Assurance	Application: ATSP and FD
Operational scenario	Distributed air-ground traffic management.	
CE 5 characteristics	Operations will include the process of distributed decision-making for its execution and execution of free maneuvering.	
Suggested Metrics: (examples)	Workload: Air traffic workload input technique (ATWIT). Situational Awareness: Situational Awareness Rating Technique (SART). Human Performance: Controller Decision Evaluation. Complexity: Aircraft mixture. Flexibility: Free flight cancellations. Efficiency: Fuel consumption. Safety: Operational errors, closest point of approach, communication delay. Capacity: Aircraft in sector, flights.	

REFERENCES

Adams, S. (1998). Practical considerations for measuring situational awareness. *Proceedings for the Third Annual Symposium and Exhibition on Situational Awareness in the Tactical Air Environment,* 157-164.

Adelman, L., Cohen, M. S., Bresnick, T. A., Chinnis, J. O., & Laskey, K. B. (1993). Real-time expert system interfaces, cognitive processes, and task performance: An empirical assessment. *Human Factors, 35*(2), 243-261.

AGARD. (1989). *Human Performance assessment methods (AGARD-AG-308)*. Neuilly-sur-Seine, France: AGARD.

Allendoerfer, K. R., & Galushka, J. (1999). *Air traffic control system baseline methodology guide.* (DOT/FAA/CT-TN99/15). Atlantic City, NJ: DOT/FAA Technical Center.

American National Standards Institute (1993). *Guide to human performance measurements* (G-035-1992). Washington, DC: American Institute of Aeronautics and Astronautics.

Ames. L. L. & George, E. J. (1993*) Revision and verification of a seven-point workload estimate scale (AFFTC-TIM-93-01)*. Edwards Air Force Base, CA: Air Force Flight Test Center.

Bateman, R. P. & Thompson, M. W. (1986). Correlation of predicted workload with actual workload measured using the subjective workload assessment technique. *Proceedings of the SAE AeroTech Conference*. (pp?) City, state: Pub.

Bergeron, H. P. (1968). Pilot response in combined control tasks. *Human Factors, 10*, 277-282.

Bisseret, A. (1970). Operational memory and structure of work. *Bulletin de Psychologie, 24,* 280-294.

Boone, J. O., Steen, J., & Van Buskirk, L. K. (1980). *System performance, error rates, and training time for persons on the radar training facility pilot position* (DOT/FAA/AM-80/5*).* Washington, DC: Office of Aviation Medicine.

Borg, C. G. (1978). Subjective aspects of physical and mental load. *Ergonomics, 21,*215-220.

Bortolussi, M. R., Kantowitz, B. H., & Hart, S. G. (1986). Measuring pilot workload in a motion base trainer: A comparison of four techniques. *Applied Ergonomics, 17*, 278-283.

Breitler, A., L, Lesko, M. J., Kirk, K. M., 1996, Effects of Sector Complexity and Controller Experience on Probability of Operational Errors in Air Route Traffic, CAN Corporation, Task 9, FAA Contract DTFA01-95-C-00002. [2]

Buckley, E. P., DeBaryshe, B. D, Hitchner, N., & Kohn P. (1983). *Methods and measurements in real time air traffic control system simulation* (DOT/FAA/CT-TN83/26). Atlantic City NJ: DOT/FAA Technical Center.

Budescu, D. V., Zwick, R.,& Rapoport, A. A. (1986). A comparison of the eigenvalue and the geometric mean procedure for ratio scaling. *Applied Psychological Measurement, 10*, 69-78.

Carretta, T. R., Perry, D. C., & Ree, M. J. (1996). Prediction of situational awareness in F-15 pilots. *International Journal of Aviation Psychology, 6*(1), 21-41.

Casali, J. G., & Wierwille, W. W. (1983). A comparison of rating scale, secondary task, physiological, and primary-task workload estimation techniques in a simulated flight task emphasizing communications load. *Human Factors, 25*, 623-642.

Cashion, P., & Lozito, S. (1999). *How short and long-term intent information affects pilot performance in a free flight environment*. Paper presented at the HCI Aero Conference, Toulouse, France.

Charlton, S. G. (1992). Establishing human factors criteria for space control systems. *Human Factors, 34,* 485-501.

Charlton, S. G. (1996). Mental workload test and evaluation. In T. G. O'Brien & S.G. Charlton (Eds.), *Handbook of human factors testing and evaluation* (pp. 181-199). Mahwah, NJ: Erlbaum.

Chatterji, G. B., Shridhar, B, 1997, Measures of Airspace Complexity, Preliminary Draft and Unpublished Work, NASA Ames Research Center. [3]

Chiles, W. D., & Alluisi E. A. (1979). On the specification of operator or occupational workload with performance-measurement methods. *Human Factors, 21(5*), 515-528.

Chiles, W. D., Jennings, A. E., & Alluisi, E. C. (1979). Measurement and scaling of workload in complex performance. *Aviation, Space, and Environmental Medicine, 5*0, 376-381.

Colle, H., Amell, J. R., Ewry, M. E., & Jenkins, M. L. (1988). Capacity equivalence curves: A double trade-off curve method for equating task performance. *Human Factors, 30* (5), 645-656.

Computer Resource Management. (1989). *National Airspace System simulation support facility (NSSF) data reduction for dependent parallel runways* (NSSF/89/002). Washington DC: Author.

Cooper, G. E., & Harper, R. P. (1969). *The use of pilot rating in the evaluation of aircraft handling qualities* (AGARD Report No. 567). London: Technical Editing and Reproduction.

Corwin, W. H., Sandry-Garza, D. L., Biferno, M. H., Boucek, G. P., Logan, A. L., Jonsson, J. E., & Metalis, S. A. (1989). Assessment of crew workload measurement methods, techniques and procedures. Volume 1 – process, methods, and results (WRDC-TR-89-7006). Wright-Patterson Air Force Base, OH.

Credeur, L., Capron, W. R., Lohr, G. W., Crawford, D. J., Tang, D. A., & Rodgers, W. G. (1993). A comparison of final approach spacing aids for terminal ATC automation. *Air Traffic Control Quarterly, 1(2*), 135-178.

Damos, D. L. (1985). The relation between the Type A behavior pattern, pacing, and subjective workload under single- and dual-task conditions. *Human Factors, 27*(6), 675-680.

Davies, A. K., Tomoszek, A., Hicks, M. R., & White, J. (1995). AWAS (Aircrew Workload Assessment System): Issues of theory, implementation, and validation. In R. Fuller, N. Johnston, & N. McDonald (Eds.), *Human factors in aviation operations. Proceedings of the 21st conference of the European Association for Aviation Psychology (EAAP) Vol. 3.*

Derrick, W. (1983). Examination of workload measures with subjective task clusters. *Proceedings of the Human Factors Society 27th Annual Meeting*, (pp. 134-138). Santa Monica, CA: Human Factors Society.

Deutsch, S. J. & Malmborg, C. J. (1982). The design of organizational performance measures for human decision making. *IEEE Transactions on Systems, Man, and Cybernetics. SMC-12* (3), 344-352.

Donnell, M. L. (1979). *An application of decision-analytic techniques to the test and evaluation of a major air system Phase III* (TR-PR-79-6-91). McLean, VA: Decisions and Designs.

Durso, F. T., Gronlund, S. D., Lewandowsky, S., & Gettys, C. F. (1991). *Cognitive factors in the use of flight progress strips: Implication for automation*. (Cognitive Processes Laboratory Report). Norman, OK: University of Oklahoma, Department of Psychology.

Empson, J. (1987). Error auditing in air traffic control. In J. A. Wise & A. Debons (Eds.), *Information systems: Failure analysis* (NATO Advanced Science Series, Vol. F32, pp. 191-198). New York: Springer-Verlag.

Endsley, M. R. (1988). Situational awareness global assessment technique (SAGAT). *Proceedings of the National Aerospace and Electronics Conference* (pp. 789-795). City, state: Pub.

Endsley, M. R. (1990). Predictive validity of an objective measure of situation awareness. *Proceedings of the Human Factors Society 34th Annual Meeting* (pp. 41-45). Santa Monica, CA: Human Factors Society. [proceedings regularly published]

Endsley, M. R. (1994). Situation awareness in dynamic human decision making: Theory. In R. D. Gilson, D. J. Garland, & J. M. Koonce (Eds.), *Situational Awareness in Complex System*s. Daytona Beach, FL: Embry Riddle Aeronautical University Press.

Endsley, M. R. (1996). Situation awareness measurement in test and evaluation. In T. G. O'Brien & S. G. Charlton (Eds.), *Handbook of Human Factors Testing and Evaluation* (pp. 159-180). Mahwah, NJ: Erlbaum.

Endsley, M. R. (1997). *Situation awareness, automation and free flight.* Cambridge, MA: Massachusetts Institute of Technology.

Endsley, M. R., Mogford, R. H., Allendoerfer, K. R., Snyder, M. D., & Stein, E. S. (1997). *Effect of free flight conditions on controller performance, workload, and situation awareness* (DOT/FAA/CT-TN97/12). William J. Hughes Technical Center, Atlantic City International Airport, NJ.

Fadden, D. (1982). *Boeing model 767 flight deck workload assessment methodology.* Paper presentation at the SAE Guidance and Control System Meeting, Williamsburg, VA.

Fairclough, S. H., Ashby, M. C., & Parkes, A. M. (1993). In-vehicle displays, visual workload and visibility evaluation. In Gales, A. G., Brown, I. D., Haslegrave, C. M., Kruysse, H. W., & Taylor, S. P. (Eds.), *Vision in Vehicles – IV*. Amsterdam: North-Holland.

Federal Aviation Administration (1998). *Air Traffic Control* (DOT/FAA/Order 7110.65L). FAA, Washington, DC.

Federal Aviation Administration & National Aviation and Space Administration (2001). *Air-Ground Integration Experiment.* Draft Report.

Fisher, S. (1975). The microstructure of dual task interaction.[no colon?] The effect of task instructions on attentional allocation and a model of attentional-switching. *Perception, 4,* 459-474.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51(4*), 327-328.

Fleming, K., Lane, J., & Corker, K. (2000). *Measuring controller reactions to free flight in a complex transition sector.* Advanced Air Transportation Technologies Office, NASA Ames Research Center, CA.

Fracker, M. L. (1989). Attention allocation in situation awareness. *Proceedings of the Human Factors Society 33rd Annual Meetin*g, (pp. 1396-1400). Santa Monica, CA: Human Factors Society.

Fracker, M. L. & Davis, S. A. (1990). Measuring operator situation awareness and mental workload. *Proceedings of the Fifth Mid-Central Ergonomics/Human Factors Conference* (pp. 23-25). Dayton, OH: pub?

Fracker, M. L. & Davis, S. A. (1991). *Explicit, implicit, and subjective rating measures of situation awareness in a monitoring task* (AL-TR-1991-0091). Wright-Patterson Air Force Base, OH.

Frankish, C., & Noyes, J. (1990). Source of human error in data entry task using speech input. *Human Factors, 32*(6), 697-716.

Galushka, J., Frederick, J., Mogford, R., & Krois, P. (1995). *Plan view display baseline research report* (DOT/FAA/CT-TN95/45). Atlantic City, NJ: DOT/FAA Technical Center.

Gartner, W. B. & Murphy, M. R. (1979). Concepts of workload. In B.O. Hartman & R. E. McKenzie (Eds.)., *Survey of Methods to Assess Workload.* AGARD-AG-246.

Gawron, V. J. (2000). *Human Performance Measures Handbook*. Mahwah, NJ: Lawrence Erlbaum.

George, E. & Hollis, S. (1991, December). *Scale Validation in Flight Test.* Edwards Air Force Base, CA: Flight Test Center.

Gopher, D. (1983). The workload book: assessment of operator workload to engineering systems (NASA-CR-1665960. Moffett Field, CA: NASA Ames Research Center.

Green, S. M., & Grace, M. P. (1999). Conflict-free planning for en route spacing - a concept for integrating conflict probe and miles-in-trail. *AIAA Guidance, Navigation, and Control Conference and Exhibit, Collection of Technical Papers,* 1 (A99-36576 09-63), 302-312.

Grosjean, V. & Terrier, P. (1999). Temporal awareness: pivotal in performance? *Ergonomics, 42*(11), 1443-1456.

Gunning, D., & Manning, M. (1980). The measurement of aircrew task loading during operational flights. *Proceedings of the Human Factors Society 24th Annual Meeting* (pp. 249-252). Santa Monica, CA: Human Factors Society.

Guttman, J. A., Stein, E., & Gromelski, S. (1995). *The influence of generic airspace on air traffic controller performance* (DOT/FAA/CT-TN95/38). Atlantic City, NJ: DOT/FAA Technical Center.

Hadley, G. A., Guttman, J. A., & Stringer, P. G. (1999). Air Traffic Control Specialist Performance Measurement Database. (DOT/FAA/CT-TN99/17). Atlantic City, NJ: DOT/FAA Technical Center.

Hamilton, D. B., Bierbaum, C. R., & Fulford, L. A. (1991). *Task Analysis/Workload (TAWL) User's guide – version 4.0* (ASI 690-330-90). Fort Rucker, AL: Anacapa Sciences.

Harris, R. L., Glover, B. J., & Spady, A. A. (1986). *Analytical techniques of pilot scanning behavior and their application (NASA Technical Paper 2525)*. Hampton VA: NASA Langley Research Center.

Harris, R. M., Hill, S. G., Lysaght, R. J., & Christ, R. E. (1992). *Handbook for Operating the OWLKNEST Technology* (ARI Research note 92-49). Alexandria, VA: United States Army Research Institute for the Behavioral and Social Sciences.

Hart, S. G., Battiste, V., & Lester, P. T. (1984). POPCORN: A supervisory control simulation for workload and performance research (NASA-CP-2341). *Proceedings of the 20th Annual Conference on Manual Control* (pp. 431-453). Washington, DC: NASA.

Hart, S. G., & Bortolussi, M. R. (1984). Pilot errors as a source of workload. *Human Factors, 25*(5), 545-556.

Hart, S. G. & Hauser, J. R. (1987). In-flight application of three pilot workload measurement techniques. *Aviation, Space and Environmental Medicine, 58,* 402-410.

Hart, S. G., & Staveland, L. E. (1987). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), *Human Mental Workload*. Amsterdam: Elsevier.

Hedge, J. W., Borman, W. C., Hanson, M. A., Carter, G. W., & Nelson, L. C. (1993). *Progress toward development of ATCS performance criterion measure* (Institute Report No. 235). Minneapolis: Personnel Decision Research Institutes.

Hicks, T. G., & Wierwille, W. W. (1979). Comparison of five mental workload assessment procedures in a moving base driving simulator. *Human Factors, 21,* 129-143.

Hilburn, B. G., Bakker, M. W., Pekela, W. D., & Parasuraman, R. (1997). The effect of free flight on air traffic controller mental workload, monitoring, and system performance. *Proceedings of the Free Flight Tenth European Aerospace Conference.*

Hopkin, V. D. (1979). Mental workload measurement in air traffic control. In N. Moray (Ed.), *Mental workload: Its theory and measurement* (pp.381-385). New York: Plenum Press.

Hopkin, V. D. (1980). The measurement of the air traffic controller. *Human Factors, 22 (5*), 547-560.

Hopkin, V. D., & Ledwith, F. (1963). *Laboratory studies in conflict detection I: Traffic density and conflicts* (RPRC/1206). Fambrough, UK: RAF Institute of Aviation Medicine.

Jensen. S. E. (1995). Developing a flight workload profile using continuous subjective assessment of workload (C-SAW). *Proceedings of the 21st Conference of the European Association for Aviation Psychology* (chapter 46).

Kahn, M. J. U., Tan, K. C., & Beaton, R. J. (1990). Reduction of cognitive workload through informational chunking. *Proceedings of the Human Factors Society 34th Annual Meeting* (pp. 1509-1513). Santa Monica, CA: Human Factors Society.

Keppel, G. (1982). *Design and analysis: A researcher's handbook.* Englewood Cliffs, NJ: Prentice-Hall, Inc.

Kilmer, K. J. , Knapp, R., Burdsal, C., Borresen, R., Bateman, R., & Malzahn, D. (1988). Techniques of subjective assessment: A comparison of the SWAT and modified Cooper-Harper scale. *Proceedings of the Human Factors Society 32nd Annual Meeting,* 155-159.

Kinney, G. C. (1977). *The human element in air traffic control: Observations and analyses of the performance of controllers and supervisors in providing ATC services* (MITRE Technical Report MTR-7655). McLean, VA: MITRE.

Knecht, W., Smith, K., and Hancock, P.A., 1996, A dynamic conflict probe and index of collision risk, Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting. [6]

Kopardekar, P. (2000). Dynamic Density – A Review of Proposed Variables**.** Report prepared for NAS Advanced Concepts Branch, ACT-540. Atlantic City, NJ: DOT/FAA Technical Center.

Kopardekar, P., and DiMeo, K, 1997, Dynamic Density Variables, PowerPoint Briefing. [1]

Kopardekar, P, Carmen, K, Sacco, N., Duley, J. D., Legan, B. M. Legan, & Helbing, K. G. (2001). Subtask 5: simulation fidelity requirements literature review and guidelines. (CTO2 Subtask 5 report) Moffett Field, CA: NASA Ames Research Center.

Kopardekar, P, and Sacco, N. (2001). Subtask 3: Human Factors Evaluations of NASA DAG-TM Concept Elements and Decision Support Tools. (CTO2 Subtask 3 report) Moffett Field, CA: NASA Ames Research Center.

Lafon-Milon, M. T. (1981). *Mental representation of vertical separation in the performance of air traffic control: Three-dimensional representation of future states* (Rapport INRIA nol. COR66). Le Chesnay, France: INRIA.

Lapan, Y.A. (1985). Spatial representation and the activity of the air traffic controller. *Vestnik Moskoskogo Universiteta Seriya 14 Psikhologiya, 14,* 68-70.

Laudeman, I.V., Brasil, C. L., and Branstrom, R. 1996, Air Traffic Control in Restructured Airspace: A Study in Tactical Decision Making, Power point presentation.

Leighbody, G., Beck, J., & Amato, T. (1992). An operational evaluation of air traffic controller workload in a simulated en route environment. *37th Annual Air Traffic Control Association Conference Proceedings* (pp. 122-130). Arlington, VA: Air Traffic Control Association.

Leplat, J., & Hoc J. (1981). Subsequent verbalization in the study of cognitive processes. *Ergonomics, 24(10*), 743-755.

Lidderdale, I. G. (1987). Measurement of aircrew workload during low-level flight, practical assessment of pilot workload (AGARD-AG-282). Proceedings of NATO Advisory Group for Aerospace Research and Development (AGARD). Neuilly-sur-Seine, France: AGARD.

Lysaght, R. J., Hill, S. G., Dick, A. O., Planmondon, B. D., Linton, P. M., Wierwille, W. W., Zaklad, A. L., Bittner, A. C., & Wherry, R. J. (1989). *Operator workload: comprehensive review and evaluation of operator workload methodologies* (Technical Report 851). Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences.

Mackintosh, M., Dunbar, M., Lozito, S., Cashion, P., McGann, A., Dulchinos, V., Ruigrok, R., Hoekstra, J., & Van Gent, R. (1998). *Self-separation from the air and ground perspective.* Paper presented at the Second USA/Europe Air Traffic Management Research and Development Seminar, Orlando, FL.

Magyarits, S. & Kopardekar, P. (2000). *Dynamic Density: The Measure of Air Traffic Control Sector Complexity.* Draft Report.

Manning, C. A., Albright, M. S., Mills, S.H., Rester, D. Rodgers, M. D., & Vardaman, J. J. (1995). *Setting Parameters for POWER (Performance and Objective Workload Evaluation Research*). Poster presentation at the 67th Annual Scientific Meeting of Aerospace Medical Association.

McDonnell, J. D. (1968). *Pilot rating techniques for the estimation and evaluation of handling qualities* (AFFDL-TR-68-76). Wright Patterson Air Force Base, OH: Air Force Flight Dynamics Laboratory.

McKenzie, R. E., Buckley, E. P., & Sarlanis, K. (1966). *An exploratory study of physiological measurements as indicators of air traffic control sector workload* (Memorandum Report). Atlantic City, NJ: National Aviation Facilities Experimental Center.

McNally, B.D., Laudeman, I.V., Mayhugh, B., 1997, Field Test Evaluation Plan for a Conflict Prediction and Conflict Resolution System, NASA Ames Research Center. [5]

Means, B., Mumaw, R., Roth, C., Shlager, M., McWilliams, E., Gagne, E., Rice, V., Rosenthal, D., & Heon, S. (1988). *ATC training analysis study: Design of the next generation ATC training system* (OPM Work Order 342-036). Richmond, VA: Human Resources Research Organization (HumRRO).

Meister, D. (1986). *Human factors testing and evaluation.* New York: Elsevier.

Meshkati, N., Hancock, P. A., & Rahimi, M. (1990). Techniques in mental workload assessment. In Wilson, J. R. and Corlett, E. N., (Eds.). *Evaluation of a human work. A practical ergonomics methodology*. New York: Taylor & Francis.

Mogford, R. H., Guttman, J. A., Morrow, S. L., & Kopardekar, P. (1995). *The complexity construct in air traffic control: A review and synthesis of the literature* (DOT/FAA/-CT TN95/22). Atlantic City, NJ: DOT/FAA Technical Center.

Mogford, R. H., Murphy, E. D., Yastrop, G., Guttman, J. A., & Roske-Hofstrand, R. J. (1993). *The application of research techniques for documenting cognitive processes in air traffic control* (Report No. DOT/FAA/CT-TN93/39). Atlantic City, NJ: Federal Aviation Administration.

Montgomery, D. C. (1991). *Design and analysis of experiments* (3rd ed.). New York, NY: John Wiley & Sons.

Mosier, K. L., & Chidester, T. R. (1991). Situation assessment and situation awareness in a team setting. In *R. M. Taylor (Ed.) Situation awareness in dynamic systems (IAM Report 708)*. Farnborough, UK: Royal Air Force Institute of Aviation Medicine.

Muckler, F. A., & Seven, S. A. (1992). Selecting performance measures: "Objective" versus "subjective" measurement. *Human Factors, 3*4, 441-456.

Muniz, E. J., Salas, E., Stout, R. J., & Bowers, C. A. (1998). The validation of a team situational awareness measure. *Proceedings for the Third Annual Symposium and Exhibition on Situational Awareness in the Tactical Air Environment*. Patuxent River, MD: Naval Air Warfare Center Aircraft Division.

Muniz, E. J., Stout, R. J., Bowers, C. A. & Salas, E. (1998). A methodology for measuring team situational awareness: Situational awareness linkied indicators adapted to novel tasks (SALIANT). *The First Annual Symposium/Business Meeting of the Human Factors and Medicine Panel on Collection Crew Performance in Complex Systems, Edenburg, United Kingdom.*

Nickels, B. J., Bobko, P., Blair, M. D., Sands, W. A., & Tartak, E. L. (1995). *Separation and Control Hiring Assessment (SACHA): Final job analysis repo*rt. Bethesda, MD: University Research Corporation.

Nieva, V. F., Fleischman, E. A., & Rieck, A. (1985). *Team dimensions: Their identity their measurement and their relationships Research Note 85-1*2. Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences.

O'Donnell, R. D., & Eggemeier, F. T. (1986). Workload assessment methodology. In K. R. Koff & J. Thomas (Eds.), *Handbook of Perception and Human Performance, 2, Cognitive processes and performance.* New York: Wiley.

O'Brien, T. G. (1996). Anthropometry, workspace, and environmental test and evaluation. In T. G. O'Brien & S. G. Charlton (Eds.), *Handbook of Human Factors Testing and Evaluatio*n, (pp. 223-264). Mahwah, NJ: Lawrence Erlbaum.

Ozmore, R. E., Morrow, S. L. (1996). *Evaluation of dual simultaneous instrument landing system approaches to runways spaced 3000 feet apart with one localizer offset using a precision runway monitor system* (DOT/FAA/CT-96/2). Atlantic City, NJ: DOT/FAA Technical Center.

Paas, F. G. W. C. & van Merrienboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors, 35*(4), 737-743.

Paul, L. (1990, October). *Using simulation to evaluate the safety of proposed ATC operations and procedures* (DOT/FAA/CT-TN90/22). Atlantic City, NJ: DOT/FAA Technical Center.

Pearson, R. G. & Byars, G. E. (1956). *The development and validation of a checklist for measuring subjective fatigue* (TR-56-115). Texas: Brooks Air Force Base, School of Aerospace Medicine.

Perez, W. A., Masline, P. J., Ramsey E. G., & Urban, K. E. (1987). *Unified tri-services cognitive performance assessment battery: Review and methodology* (AAMRL-TR-87-007). Wright-Patterson Air Force Base, OH: Aerospace Medical Research Laboratory.

Philips, C. T. (2000). *Detailed description for CE-5: En route free maneuvering* (NAS2-98005 RTO-41). Moffett Field, CA: Technical Research in Advanced Air Transportation Technologies.

Pollock, V., Cho, D. W., Rekar, D., & Volavka, J. (1979). Profile of mood states: The factors and their correlates. *Journal of Nervous Mental Disorders. 167,* 612-614.

Porterfield, D. H. (1997). Evaluating controller communication time as a measure of workload. *The International Journal of Aviation Psychology, 7*(2), 171-182.

Redi, G. B., & Nygren, T. E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P.A. Hancock & N. Mehtaki (Eds.), *Human mental workload* (pp. 185-218). Amsterdam: Holland.

Roenker, D. L., Thompson, C. P., & Brown, S. C. (1971). Comparison of measures for the estimation of clustering in free recall. *Psychological Bulletin, 76,* 45-48.

Rolfe, J. M. (1971). The secondary task as a measure of mental load. In W. T. Singleton, J. G. Fox, & D. Whitfield (Eds.), *Measurement of man at work.* London: Taylor & Francis.

Roscoe, A. H. (1984). Assessing pilot workload in flight. Flight test techniques. *Proceedings of NATO Advisory Group for Aerospace Research and Development* (AGARD-CP-373). Neuilly-sur-Seine: France.

Roske-Hofstrand, R. J. (1989). Video in applied cognitive research for human-centered design. *SIGCHI Bulletin, 21(2*), 75-77.

Saaty T. L. (1980). *The analytical hierarchy proces*s. New York: McGraw-Hill.

Sanders, M. S. & McCormick, E. J. (1993). *Human Factors in Engineering and Design* (7th ed.) New York< NY: McGraw-Hill Inc.

Sanderson, P. (1990). Verbal protocol analysis in three experimental domains using SHAPA. *Proceedings of the Human Factors Society 34th Annual Meeting,* (pp. 1280-1284). Santa Monica, CA: Human Factors Society.

Sarter, N. B. & Woods, D. D. (1991). Situational awareness: a critical but ill-defined phenomenon. *International Journal of Aviation Psychology, 1*(1), 45-57.

Selcon, S. J., & Taylor, R. M. (1989). *Evaluation of the situational awareness rating technique (SART) as a tool for aircrew systems design*. AGARD Conference Proceedings No. 478, Neuilly-sur-Seine: France.

Selcon, S. J., & Taylor, R. M. (1991). Decision support and situational awareness. In R. M. Taylor (Ed.), *Situational Awareness in Dynamic Systems* (Rep. No. IAM-708). Farnborough, UK: Royal Air Force Institute of Aviation Medicine.

Shachem, A. A. (1983). A shortened version of the profile of mood states. *Journal of Personality Assessment, 47,* 305-306.

Shively, R. J., Brickner, M., & Silbiger, J. (1997). A computational model of situational awareness instantiated in MIDAS. http://caffeine.arc.nasa.gov/midas/Tech_Reports.html.

Slocum, G. K., Williges, B. H., & Roscoe, S. N. (1971). Meaningful shape coding for aircraft switch knobs. *Aviation Research Monographs, 1,* (3), 27-40.

Sollenberger, R. L., Stein, E. S., & Gromelski, S. (1997). *The development and evaluation of a behaviorally based rating form for the assessment of air traffic controller performance* (DOT/FAA/CT-TN96-16). Atlantic City, NJ: DOT/FAA Technical Center.

Sorensen, J. A. (2000). *Detailed description for CE11. Terminal arrival: Self-spacing for merging and in-trail separation* (NAS2-98005 RTO-41). Moffett Field, CA: Technical Research in Advanced Air Transportation Technologies.

Speyer, J., Fort, A., Fouillot, J., & Bloomberg, R. (1987). Assessing pilot workload for minimum crew certification. In A. H. Roscoe (Ed.), *The practical assessment of pilot workload* (AGARD graph No, 282, pp. 90-115). Neuilly-sur-Seine: France.

Stein, E. S. (1984). *The measurement of pilot performance: A master-journeyman approach* (DOT/FAA/CT-83/15). Atlantic City, NJ: DOT/FAA Technical Center.

Taylor, R. M. (1989). Situational awareness rating technique (SART): The development of a tool for aircrew systems design. *Proceedings of the NATO Advisory Group for Aerospace Research and Development* (AGARD-CP-478).

Taylor, R. M. (1990). Situational awareness: aircrew constructs for subject estimation (IAM-R670).

Taylor, R. M., & Selcon, S. J. (1991). *Subjective measurement of situational awareness.* In R. M. Taylor (Ed.) situational awareness in dynamic systems (IAM Report 708). Farnborough, UK: Royal Air Force Institute of Aviation Medicine.

Taylor, R. M., Selcon, S. J., & Swinden, A. D. (1995). Measurement of situational awareness and performance: A unitary SART index predicts performance on a simulated ATC task. *Proceedings of the 21st Conference of the European Association for Aviation Psychology, Chapter 41.*

Vidulich, M. A. (1989a). The use of judgment matrices in subjective workload assessment: The subjective workload dominance (SWORD) technique. *Proceedings of the Human Factors Society 33rd Annual Meeting* (pp.1406-1410).

Vidulich, M. A. (1989b). The objective measures of workload: Should a secondary task be secondary?. *Proceedings of the Fifth International Symposium on Aviation Psychology* (pp.802-807).

Vidulich, M. A. (1991). The Bedford scale: does it measure spare capacity? *Proceedings of the 6th International Symposium on Aviation Psychology, 2,* 1136-1141.

Vortac, O. U., Edwards, M. B., & Manning, C. A. (1995). *Functions of external cues in prospective memory* (DOT/FAA/AM-95/9). Washington, DC: Office of Aviation Medicine.

Vortac, O. U., Edwards, M. B., Fuller, D. K., & Manning, C. A. (1994). *Automation and cognition in Air Traffic Control: An empirical investigation* (DOT/FAA/AM-94/3). Washington, DC: Office of Aviation Medicine.

Wickens, C. D. (1984). *Engineering Psychology and Human Performance*. Columbus, OH: Charles E. Merrill.

Wickens, C. D., Mountford, S. J., & Schreiner, W. Multiple resources, task-hemispheric integrity, and the differences in time-sharing. *Human Factors, 23*, 211-229.

Wierwille, W. W., Casali, J. G., Connor, J. G., & Rahimi, M. (1985). Evaluation of the sensitivity and intrusion of mental workload estimation techniques. In W. Roner (Ed.) Advances in Man-Machine Systems Research. (Vol. 2, pp. 51-127). Greenwich, CT: J. A. I. Press.

Wierwille, W. W. & Conner, S. A. (1983). Evaluation of 20 workload measures using a psychomotor task in a moving-base aircraft simulator. *Human Factors, 25*(1), 1-16.

Wierwille, W. W. & Eggemeier, F. T. (1993). Recommendations for mental workload measurement in a test and evaluation environment. *Human Factors, 35*(2), 263-281.

Wierwille, W. W., Rahimi, M., & Casali, J. G. (1985). Evaluation of 16 measures of mental workload using a simulated flight task emphasizing mediational activity. *Human Factors, 27*(5), 489-502.

Wierwille, W. W., Williges, R. C., & Schiflet, S. G. (1979). Aircrew workload assessment techniques. In B. O. Hartmand & R. E. McKenzie (Eds), *Survey of methods to assess workload.* AGARD-AG-246.

Willems, B., Allen, R. C., & Stein, E. S. (1999). *Air traffic control specialist visual scanning II: Task load, visual noise, and intrusions into controlled airspace* (DOT/FAA/CT-TN99/23). Atlantic City, NJ: DOT/FAA Technical Center.

Wing, D. J., Adams, R. J., Duley, J. A., Legan, B. M., Baremore, B. E., & Moses, D. (in press). Airborne use of traffic intent information in a distributed air-ground traffic management concept: Experiment design and preliminary results. NASA Technical Memorandum publication.

Wolf, J. D. (1978). *Crew Workload Assessment: Development of a Measure of Operator Workload (AFFDL-TR-78-165).* Wright-Patterson AFB, OH: Air Force Flight Dynamics Laboratory.

Zaklad, A., Zachary, W., & Davis, D. A. (1987). A cognitive model of multisensor correlation in an advanced aircraft environment. *Proceedings of the Fourth Midcentral Ergonomics/Human Factors Conference,* Urbana, IL.

APPENDIX A – LITERATURE REVIEW METRIC DATABASE

A.1 LITERATURE REVIEW BACKGROUND

The database presented in Table A1 contains the name of the metric with a basic definition or description. In some cases, the metric need has been identified, but requires further development. Included with each definition and description are relevant studies that have developed, validated, reviewed, or utilized the metric. Each metric is classified according to human performance, and overall air traffic system level constructs. Human performance constructs include: communications, conflict, error, usability, human performance, situational awareness, and workload measures. Air traffic system level constructs refer to system or air traffic capacity, complexity, efficiency, flexibility, safety, and simulation fidelity. Metrics may apply to one or several areas of the air traffic management system. In order to effectively measure aspects of DAG-TM concepts, it is essential to identify metrics that can be transferred across all decision-making agents. Therefore, this database also categorizes metrics that apply across the ATSP, FC, and the AOC.

This Page Intentionally Left Blank

Table A1. Human Performance and System Metrics

		Human Performance Metrics								System Level Metrics						Agents		
METRIC	**DEFINITION/DESCRIPTION**	**COMMUNICATIONS**	**CONFLICT**	**ERROR**	**DECISION MAKING**	**SITUATIONAL AWARENESS**	**USABILITY**	**WORKLOAD**	**OTHER**	**CAPACITY**	**COMPLEXITY**	**EFFICIENCY**	**FLEXIBILITY**	**SAFETY**	**SIMULATION FIDELITY**	**ATSP**	**FC**	**AOC**
Acceptance of separation role-Pilot/Controller	Metric needs further definition. Frequency of pilot requested FF cancellations, frequency of controller cancellations (FAA & NASA, 2001).												X			X	X	
Accessibility of controls and flight strips	Measure of the usability of the flight strips and accessibility of the flight strips bay (Allendoerfer & Galushka, 1999; Durso, Gronlund, Lewandowsky & Gettys, 1991).						X	X								X		
Acquisitions	Number of times aircraft acquired localizer during experimental run (Ozmore & Morrow, 1996).							X		X	X	X				X	X	

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Adequacy of radar and radio coverage	Incomplete radar or radar coverage causes additional complexity due to lack of automated aids available with radar and the need to relay information from aircraft in radio coverage to aircraft not directly accessible (Mogford, Guttman, Morrow & Kopardekar, 1995).										X					X		
Air traffic workload input technique	Subjective workload measured at standard intervals during the simulation (FAA & NASA 2001; Sollenberger, Stein & Gromelski, 1997).							X								X		
Aircraft – control count	Interval count and total number of aircraft under control for duration of experimental run (Magyarits & Kopardekar, 2000).							X		X	X					X	X	X
Aircraft mixture	In the AT mix, greater variety of slow and fast aircraft brings greater complexity due to potential overtaking conflicts (Mogford et al., 1995).									X	X					X	X	
Aircraft Pair Inter-Arrival Error	The difference between arrival errors of sequential arrival aircraft, in terms of aircrafts' actual and scheduled times of arrival (Credeur et al., 1993).			X		X								X		X	X	X
Aircraft path changes	Frequency of aircraft changes heading, speed or altitude (Ozmore & Morrow, 1996).							X								X	X	

Raytheon

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Aircraft proximity index variable	API is a weighted measure of conflict intensity where 100 is a mid-air collision and 1 is a minor violation of separation standards (FAA & NASA, 2001; Manning et al., 1995; Ozmore & Morrow, 1996).		X			X								X		X	X	
Aircraft separation – Average distance	Average slant range of an aircraft pair (in feet) (Paul, 1990).		X			X		X						X		X		
Aircraft -Time in sector	Average time an aircraft spent under a controller's control (Buckley et al. 1983; FAA & NASA, 2001).							X			X					X		
Aircrew Workload Assessment System	Timeline analysis software developed by Boeing to predict workload. Requires 3 inputs: 1) second by second description of pilot tasks during flight; 2) demands on each of Wicken's (1981) multiple resource theory processing channels; 3) effects of simultaneous demand on a single channel (Davies, Tomoszek, Hicks & White, 1995).				X			X									X	
Airline Hubbing	Airline hubbing causes more complexity by bringing in more aircraft from the same company with similar call signs, and more aircraft are arriving and departing on fewer airways (Mogford et al., 1995).				X						X							X
Airspace conflict frequency cumulative durations variable	Duration of intrusion into restricted airspace (FAA & NASA, 2001; Manning et al., 1995; Ozmore & Morrow, 1996).		X			X								X		X	X	

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Airspace conflict frequency variable	Frequency of intrusion into restricted airspace (FAA & NASA, 2001; Manning et al., 1995; Ozmore & Morrow, 1996).		X									X		X			X	
Altitude assignments	Extent to which controller correctly assigns altitudes to aircraft under his/her control (Galushka et al, 1995).				X			X				X				X		
Altitude changes	Frequency of FC initiated altitude changes (Allendoerfer & Galushka, 1999).				X			X								X	X	
Altitude clearances	Frequency of altitude clearances issued during a run (CRM, 1989; Ozmore & Morrow, 1996; Sollenberger, Stein & Gromelski, 1997).				X						X					X		
Altitude filtering	FC preferred CDTI altitude filter settings (FAA & NASA, 2001).		X		X			X				X				X	X	
Analytical Hierarchy Process	This process uses paired comparisons to measure workload. Subjects rate which of a pair of conditions has the higher workload. All combinations of conditions must be compared (Saaty, 1980).				X		X	X								X	X	X
Arrival delay	Average time aircraft arrive after scheduled arrival time.								X	X						X	X	X
Arrivals	Number of landings occurring during experimental run; time between arrivals, spacing between arrivals (Allendoerfer & Galushka, 1999).							X		X	X					X	X	

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Bedford Workload Scale	Roscoe (1984) described a modification of the Cooper-Harper scale created by trial and error with test pilots at Royal Aircraft Establishment, Bedford England. Bedford scale retains binary decision tree of Cooper Harper Scale (Vidulich, 1991; Lidderdale, 1987).				X			X									X	
Behaviorally Anchored Expert Observations	Expert observers rate various performance dimensions. Rating performance of specific observable controller actions reduces need for observers to make unreliable inferences about controller performance (Sollenberger, Stein & Gromelski, 1997).				X				X							X		
Blundering aircraft and next aircraft receiving a path change message	Planned deviations from localizer where one aircraft crosses another's landing path (Ozmore & Morrow, 1996; Wing et al., in press).			X	X	X								X		X	X	
Blunders and associated conflicts	An unexpected turn by an aircraft already established on localizer toward another aircraft on adjacent approach (Ozmore & Morrow, 1996; Wing et al., in press).			X	X	X		X				X	X	X		X		
Cancel flight: Instantaneous Aircraft Count	Number of cancelled flights that occurred during an experimental run (Buckley et al., 1983).							X								X	X	

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Categorization	This method assumes a fundamental difference in the ways novices and experts classify problems. May be useful in discriminating between levels of operator competence and experience (Nickels, Bobko, Blair, Sands & Tartak, 1995).								X							X	X	X
Charlton's Measures of Human Performance in Space Control Systems	Charlton's measures of prediction are divided into 3 phases (pre-pass, contact execution, and contact termination) and 3 crew positions (ground controller, mission controller and planner analyst) (Charlton, 1992; 1996).								X							X		
China Lake Situational Awareness	A five point (based on Bedford Workload Scale) subjective rating scale designed to measure SA in flight (Adams, 1998).					X											X	
Clearance: Instantaneous Aircraft Count	Number of clearances issued during an experimental session (Buckley et al., 1983).							X				X				X	X	X
Closest-point-of-approach	Slant range of aircraft pair in conflict measured in feet (FAA & NASA, 2001; Ozmore & Morrow, 1996).		X			X								X		X	X	

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Clustering	Clustering refers to the degree to which a participant performs actions, typically performed consecutively, in a consecutive manner. Organized, systematic behavior is (expected to be) characteristic of well thought out behavior (Roenker, Thompson & Brown, 1971; Kahn, Tan & Beaton, 1990).				X			X								X	X	
Clutter	Amount of clutter in a given area of user's screen (Wing et al., in press).					X	X	X								X	X	X
Communication - Amount of coordination	Coordination requires communication with ground controllers and imposes additional task load due to point outs and waiting for the coordinating sector to approve or disapprove (Mogford et al., 1995).	X						X		X			X			X	X	
Communication - FC coordination	Frequency of communication between FC.	X						X				X	X	X			X	
Communication - Frequency congestion	Total frequency, duration and type (i.e. air to ground, ground to air, etc.) of overall communications on frequency (Buckley et al., 1983; FAA & NASA 2001).	X						X								X	X	
Communication delay	The accumulated time variable based on durations of time between the aircraft calls for service and controllers' initial response (Ozmore & Morrow, 1996).	X			X			X				X		X		X	X	

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Communication efficiency	The extent to which a controller can handle communication tasks (Galushka et al., 1995).	X										X				X		
Communication - Frequency congestion	Adds to complexity due to increased difficulty in communicating with several aircraft on same frequency (Mogford et al., 1995).	X									X					X	X	
Complex aircraft routings	Complex aircraft routings require more attention to aircraft due to crossing points, turns and potential conflicts with other aircraft. Ideally controllers send aircraft direct to a fix outside sector (Mogford et al., 1995).										X					X		
Complexity measures activity variance	Measure of aircraft clustering within user specifiable criteria (e.g., 10 mi.). Higher index means more aircraft are clustering and are more likely to conflict (Sollenberger, Stein & Gromelski, 1997).										X					X		
Computerized Rapid Analysis of Workload	CRAWL is a computer program that helps designers predict workload in systems. Inputs are mission timelines and task descriptions. Tasks are described in terms of cognitive, psychomotor, auditory and visual demands (Vortac, Edwards, Fuller & Manning, 1994).							X									X	

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Confidence in automation-FC/Controller/AOC	Metric needs further definition. Subjective questionnaire, debriefing, heads down time (pilot), frequency of confirmation communications (Wing et al., in press).				X							X				X	X	X
Conflict alerts	The number of tool specific conflict alerts that occur during simulation (FAA & NASA 2001; Sollenberger, Stein & Gromelski, 1997; Hopkin & Ledwith, 1963).		X			X								X		X	X	
Conflict detection and resolution action - delta time	Duration of time from conflict detection to initiation of conflict resolution plan (Wing et al., in press).		X		X							X		X		X	X	
Conflict detection prior to automation alert	Frequency and average time of human conflict detection prior to automated alert (FAA & NASA 2001; Wing et al., in press; Hilburn, Bakker, Pekela & Parasuraman, 1997).		X		X			X		X		X		X		X	X	X
Conflict frequency cumulative durations variable, – between-sector	Duration of conflict between aircraft pair when each has a different controller (FAA & NASA, 2001; Manning et al., 1995; Ozmore & Morrow, 1996).		X			X		X						X		X		
Conflict frequency variable, between-sector	Number of conflicts between aircraft pair when each has a different controller (FAA & NASA, 2001; Manning et al., 1995; Ozmore & Morrow, 1996).		X											X		X	X	
Conflict resolution strategy	Type of maneuver (vector, altitude, speed) planned for conflict resolution (FAA & NASA, 2001).		X		X	X								X		X	X	

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Conflicts - Horizontal separation (miles)	Horizontal separation of aircraft pair in conflict and is measured in miles (Buckley et al., 1983).		X			X								X		X	X	
Conflicts - relationship of ILS	Relationships of ILS conflicts are: X-1: side-by-side, X-2: ILS between, X-3: two ILS's between (Buckley et al., 1983).		X											X		X	X	
Continuous Subjective Assessment of Workload	C-SAW requires subjects to provide 1-10 ratings (Bedford Scale descriptors) while viewing a videotape of their flight immediately after landing (Jensen, 1995).				X			X				X					X	
Control Input Activity	Control input activity for the wheel and column as a measure of flight-path control (Corwin et al., 1989).								X			X					X	
Control Movements/Unit Time	The number of control inputs made (summed over each control used by one operator) divided by the unit of time over which the measurements were made (Wierwille & Conner, 1983).				X			X									X	
Controller Decision Evaluation	The method presents a traffic situation unfolding in a film/video and requires controller to determine next appropriate action (Borg, 1978; Kinney, 1977).				X				X							X		
Controller Keystrokes - Communication Activity	Number of keystrokes entered at controller's keyboard (Porterfield, 1997; Wierwille & Connor, 1983).							X								X		

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Cooper-Harper Rating Scale	This decision tree uses adequacy of the task, aircraft characteristics, and demands on the pilot to rate handling qualities of an aircraft (Cooper & Harper, 1969).				X			X									X	
Coordination - Between-sector	Measure of taskload generated by coordination between radar and data controllers in a different sector (FAA & NASA 2001; Galushka et al., 1995; Sollenberger, Stein & Gromelski, 1997).	X			X			X				X				X	X	
Crew Situational Awareness	Expert observers rate crew coordination performance, identify and rate performance errors, then develop information transfer matrices identifying time and source of item requests and verbalized responses (Mosier & Chidester, 1991).					X											X	
Crew Status Survey	20 statements describing fatigue status (Ames & George, 1993; Pearson & Byars, 1956; George & Hollis, 1991).							X									X	
Critical Incident Technique	The CIT uses a set of procedures to collect direct observations of controller behavior to learn about planning, decision-making, and problem solving (Flanagan, 1954).				X				X							X		

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Critical Incidents Interviews	The technique consists of two interviews: the first identifies unusual or difficult situations encountered by participants; a follow-up session reviews incident descriptions and elicits possible alternatives to each action (Flanagan, 1954).				X				X							X	X	X
Data block offset	Number of times data blocks are offset during an experimental run (Nieva, Fleischman & Rieck, 1985).							X								X		
Data entry efficiency	Extent to which a controller can handle data entry tasks (Galushka et al., 1995).											X				X		
Deferred pilot requests	Number of pilot requests deferred to a later time or subsequent sector during positive ATC control operations.	X			X			X					X			X	X	
Deliberate pilot noncompliance or miscompliance - Simulation Conditions	Scenario variable where simulation pilots may not follow clearances accurately or may make path changes without a clearance (Wing et al., in press).			X	X	X								X		X	X	
Denied pilot requests	Number of pilot requests denied during positive ATC control operations.	X			X			X		X		X				X		
Departures	Number of departures occurring during an experimental run (Boone et al., 1980).										X					X		

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Deutsch and Malmborg Measurement Instrument Matrix	Measures the complexity and interaction of activities performed by organizations (Deutsch & Malmborg, 1982).				X				X							X	X	X
Deviation (feet, L-left, R-right), MX (max. deviation in feet)	Deviation from ILS enter line (in feet) (Buckley et al., 1983).			X	X	X								X		X	X	
Deviations from user preferred routes	Frequency, delta flight path, delta time for deviations (Wing et al., in press).				X			X				X	X			X	X	X
Distance aircraft under control	Distance (in miles) an aircraft flew in a simulation (Sollenberger, Stein & Gromelski, 1997).					X		X								X	X	
Domain Knowledge Test	Used to determine whether one interface design is superior to others in helping system operators acquire domain knowledge.								X							X		
D-side data entries	Extent to which data controller enters data quickly and accurately (Galushka et al., 1995).							X								X		
D-side data entry errors	Number of data entry errors accumulated by data controller (Galushka et al., 1995).			X		X								X		X		

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Dual Coding tasks	These tasks require participants to compare perceived and imagined objects, compare symbols, make mental transformations, and perform computations based on representational structures (AGARD, 1989; Fisher, 1975).					X		X	X							X	X	X
Dynamic Density	Measure of sector complexity. Takes into account number of aircraft, volume of airspace, traffic flows, weather, etc. Note several proposed metrics are currently being validated (Magyarits & Kopardekar, 2000).										X					X	X	X
Dynamic Workload Scale	This seven-point scale developed as a tool for aircraft certification has been used extensively by Airbus Industries (Speyer, Fort, Fouillot & Bloomberg, 1987).							X									X	
Enhanced Video Recordings	Reported in a paper on the use of combined video and eye movement recordings (Roske-Hofstrand, 1989).								X							X	X	
Entry into NTZ	Time an aircraft enters no transgression zone (FAA, 1998).			X	X	X								X		X	X	
Environmental factors	Measure of impact of environmental factors, e.g., workspace lighting and anthropometry on usability (Allendoerfer & Galushka, 1999; O'Brien, 1996).						X	X								X	X	X

Raytheon

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Equal Appearing Intervals	Participants rate workload in one of several categories using the assumption that each category is equidistant from adjacent categories (Hicks & Wierwille, 1979).							X								X	X	X
Exit from NTZ	Time that an aircraft leaves no transgression zone (FAA, 1998).			X	X	X		X						X		X		
Expert observers-performance ratings	Over-the-shoulder ratings of various performance dimensions by subject matter experts (Allendoerfer & Galushka, 1999).					X		X				X				X	X	
False alerts-frequency	A false alert is an automated alert for which neither aircraft has deviated from its flight plan and no separation violation occurs for the duration of the alert (Hilburn, Bakker, Pekela, & Parasuraman, 1997; Wing et al., in press).		X			X								X		X	X	
FF cancellation requests by FD	Frequency of pilot initiated free flight cancellations (FAA & NASA, 2001).				X			X				X	X	X			X	
FF cancellations by ATC	Frequency of ATC initiated free flight cancellations (FAA & NASA, 2001).				X			X				X	X	X		X	X	X
Flight Workload Questionnaire	The flight workload questionnaire is a four item behaviorally anchored rating scale. The items of the rating scale are workload category, fraction of time busy, how hard had to think, and how felt (relaxed to very stressful) (Stein, 1984).							X									X	

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Flights	Number of flights accumulated during an experimental run (FAA & NASA 2001; Magyarits & Kopardekar, 2000).							X		X	X					X	X	X
Frequency transfers - Instantaneous Aircraft Count	Number of frequency transfers that occurred during an experimental run (Buckley, et al., 1983).	X						X			X	X				X		
Fuel consumption	Fuel used by each aircraft in an experimental run for a standard distance (Wing et al., in press; Galushka et al., 1995).											X				X	X	
Glance Duration and Frequency	Duration and frequency of glances to visual displays. The longer the durations and or the greater the frequency of glances, the higher the workload (Credeur, et al., 1993; Fairclough, Ashby, & Parkes, 1993; Harris, Glover, & Spady, 1986; Willems, Allen & Stein, 1999).				X		X	X									X	
Hand off delay time (initiate to acknowledge)	Delay time from when the aircraft was handed off to when the participant controller accepted the hand-off (Manning et al., 1995).								X							X		
Hand off to/from subject	Number of hand-offs made and received by the participant during an experimental run (Manning et al., 1995).								X							X		
Handoff errors - Non Conflict Errors	Frequency with which the aircraft was handed off to the wrong controller (Manning et al., 1995).			X		X		X	X					X		X		

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Handoff misses - Non Conflict Errors	Frequency in which the aircraft crossed the sector boundary before handoff (Manning et al., 1995).			X				X				X		X		X	X	
Hart and Bortolussi Rating Scale	A single rating scale (1 to 100) to estimate workload (Hart & Bortolussi, 1984).							X									X	
Hart and Hauser Rating Scale	Hart & Hauser use a 6-item rating to measure workload during a 9-hour flight: stress, mental / sensory effort, fatigue, time pressure, overall workload and performance (Hart & Hauser, 1987).							X								X	X	X
Heading -	Frequency of heading clearances issued during a run (CRM, 1989; Ozmore & Morrow, 1996; Sollenberger, Stein, & Gromelski, 1997).				X						X					X		
Heading – FC initiated.	Number of times the aircraft changed heading (FAA & NASA, 2001).				X							X				X	X	
History trail	Number of times history trails were used during an experimental run (Nieva, Fleischman & Rieck, 1985).				X									X		X		
Hold messages	Number of hold clearances issued during an experimental run (CRM, 1989).				X			X		X		X				X		
Honeywell Cooper-Harper Rating Scale	Rating scale that uses a decision-tree structure for assessing overall task workload (Wolf, 1978).				X			X									X	
Horizontal distance (nm)	Horizontal component of slant range, (measured in nm) (Paul, 1990).		X											X		X		

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Human Interface Rating and Evaluation System	HiRes System is a generic judgment-scaling technique used to evaluate SA. Ratings are scaled to sum to 1.0 across conditions (Budescu, Zwick & Rapoport, 1986; Fracker & Davis, 1990).					X										X	X	X
Information transfer	Timeliness of information sharing between authorities.	X			X									X		X	X	X
Information, clearances, reports, beacon, miscellaneous	Number of miscellaneous clearances issued during an experimental run (CRM, 1989).							X								X		
Intersecting flight paths	Number of jet routes or victor airways that cross within the sector. The greater the occurrence the more stringent the requirement for spacing and sequencing, including vertical separation to avoid conflicts at these crossing points (Mogford, Guttman, Morrow, & Kopardekar, 1995).										X					X		
In-track time spent inside the final approach fix	Amount of controller monitoring inside final approach fix. Considered critical because of separation compression normally occurring in vicinity of outer marker (Credeur et al., 1993).								X			X				X	X	
Keyboard entry errors - Pilot	Every backspace is counted, and if a CLR key is struck, every key in that message is counted as an error (Buckley et al., 1983).			X		X		X				X		X		X		

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Keyboard messages - Pilot	Completed pilot keyboard messages (CRM, 1989).							X									X	
Keystrokes-Pilot	Number of keystrokes entered at the simulation pilot's keyboard (Davies, Tomoszek, Hicks & White, 1995; Guttman, Stein & Gromelski, 1995; Wierwille & Connor, 1983).							X								X		
Load Stress	Stress produced by increasing number of signal sources that must be attended during task (Chiles & Alluisi, 1979).							X									X	
Locus of control	Under distributed management-length of time each authority has positive control (FAA & NASA, 2001).				X	X		X				X		X		X	X	X
Magnitude Estimation	Participants are required to estimate workload numerically in relation to a standard (Borg, 1978; Hart & Staveland, 1987).							X								X	X	X
Maneuvers required to resolve a conflict	Frequency and type of maneuvers performed to resolve a potential conflict (FAA & NASA, 2001).				X			X				X				X	X	
McDonnell Rating Scale	This 10-point scale requires a pilot to rate workload based on the attentional demands of a task (McDonnell, 1968).							X								X	X	X

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Military flights	Military flights may require special handling that imposes additional taskload. They often make special requests, do not conform to procedures, and fly in formations, which they may break during flight (Mogford et al., 1995).							X			X					X		
Missed alerts-frequency	Missed alert is an automated alert for which there is no deviation from either flight plan; a loss of separation occurs before alert is issued (Hilburn, Bakker, Pekela, & Parasuraman, 1997; Wing et al., in press).		X											X		X	X	
Missed approaches - Non Conflict Errors	Frequency of missed approaches executed during a run (Buckley et al., 1983).			X	X			X									X	
Mission Operability Assessment Technique	The technique uses two 4-point rating scales, one for workload and one for technical effectiveness. Participants rate both categories for each subsystem identified in a task analysis (Donnell, 1979).				X			X				X					X	
Modified Cooper-Harper Rating Scale	This scale was developed to increase the range of applicability to situations commonly found in modern systems (Casali & Wierwille, 1983; Kilmer et al., 1988).							X									X	

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Multi-Descriptor Scale	Scale composed of six descriptors: attentional demand, error level, difficulty, task complexity, mental workload, and stress level. Descriptor is rated after task. Score is average of the six ratings (Wierwille, Rahimi, & Casali, 1985).							X								X	X	X
Multidimensional Rating Scale	Eight bipolar scales, on which subjects draw horizontal lines to indicate rating (Damos, 1985).							X								X	X	X
Multidimensional scaling	Multidimensional scaling was used for direct and indirect reconstruction of cognitive maps; diagnostic version used to study mental rotation of three-dimensional objects (Lapan, 1985).								X							X		
NASA Bipolar Rating Scale	The scale has ten subscales. Any scale not relevant to a task is given a weight of zero. A weighting procedure is used to enhance intrasubject reliability (Bortolussi, Kantowitz & Hart, 1986; Hart, Battiste & Lester, 1984).							X								X	X	X
NASA Task Load Index	TLX is a multi-dimensional subjective workload rating technique. Workload is defined as the cost incurred by human operators to achieve a specific level of performance (Hart & Staveland, 1987).							X								X	X	X

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Nieva, Fleishman, and Rieck's Team Dimensions	Authors defined five measures of team performance: (1) matching number resources to task requirements, (2) response coordination, (3) activity pacing, (4) priority assignment among tasks, and (5) load balancing (Nieva, Fleischman & Rieck, 1985).								X							X	X	X
Operational errors - Safety	An operational error is one in which separation standards were violated (FAA & NASA, 2001; Sollenberger, Stein & Gromelski, 1997).		X	X										X		X	X	
Over flights	Number of over flight aircraft in a scenario (Allendoerfer & Galushka, 1999).										X				X			
Overall Workload Scale	Bipolar scale (low to high) requiring subjects to provide single workload rating on a horizontal line divided in 20 equal intervals (Harris, Hill, Lysaght & Christ, 1992).							X								X	X	X
Parallel conflict frequency cumulative durations variable	Duration of conflict for aircraft pair conflicting on simultaneous parallel approach (CRM, 1989; Ozmore & Morrow, 1996).		X											X		X	X	
Parallel conflict frequency variable	Frequency of conflicts between aircraft on simultaneous parallel approaches (CRM, 1989; Ozmore & Morrow, 1996).		X											X		X	X	
Path change/data link messages	Number of altitude, heading or speed changes issued by controller during experimental run (Ozmore & Morrow, 1996; Sollenberger, Stein & Gromelski, 1997).				X			X								X	X	

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Performance measure guidelines	Selecting objective vs. subjective performance measures (Muckler & Seven, 1992).								X							X	X	X
Pilot Objective/Subjective Workload Assessment Technique	POSWAT is a ten point subjective scale developed at the FAA Technical Center. This modified Cooper-Harper scale does not include the binary decision tree (Stein, 1984).							X				X				X	X	X
Pilot Performance Index	Performance variable and associated performance criteria for an air transport mission to distinguish between experienced and novice pilots (Stein, 1984).								X			X					X	
Pilot RAT tool to make a request	Frequency of pilot use of route assistance tool.				X			X				X	X			X	X	
Pilot requests/messages	Total number of pilot requests for deviations, altitude changes, weather, and turbulence reports.	X			X			X				X				X	X	
Pilot Subjective Evaluation	The PSE workload scale was developed by Boeing to certifyB-767 aircraft (Fadden, 1982).							X									X	
Pilot Subjective Evaluation	The PSE scale and its companion questionnaire are completed with reference to an existing aircraft (Lysaght et al., 1989).							X						X			X	
Point-Outs	Number of between sector point outs (FAA & NASA, 2001).					X		X								X		

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Primary conflict measure for aircraft on final approaches and in trail of one another	Measures longitudinal conflicts of aircraft on approach (Green & Grace, 1999; Vidulich, 1989a).		X											X		X	X	
Profile of Mood States	The short version of PMS scale measures self-rated tension, anger, depression, vigor, fatigue, and confusion (Pollock, Cho, Rekar, & Volavka, 1979; Shachem, 1983).							X								X	X	
Quality of service - Performance	ATC services (Galushka et al., 1995).								X			X				X		
Range setting	FC preferred CDTI look ahead range in nautical miles (FAA & NASA, 2001).		X		X		X							X		X	X	
Rate of Gain of Information	Based on Hicks' law, stating reaction time is a linear function of the amount of information transmitted (Chiles & Alluisi, 1979).								X							X	X	X
R-Data entries - Performance	Extent to which radar controller enters data quickly and accurately (Galushka et al., 1995).							X								X		
R-Data entry errors - Performance	Number of data entry errors accumulated by radar controller (Galushka et al., 1995).			X		X		X								X	X	
Recall tasks	ATC researchers use recall tasks to study memory (Bisseret, 1970).								X							X	X	X

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Relative Comparison Technique	This technique draws upon an aircrew's expertise with a similar system. Relative data are collected by comparing similar items (Vortac, Edwards, & Manning, 1995).					X		X									X	
Relative Condition Efficiency	Combined ratings of workload with task performance measures to calculate relative condition efficiency (Paas & van Merrienboer, 1993).							X	X			X						
Report messages - Instantaneous Aircraft Count	Number of report messages that occurred during experimental run (Buckley et al., 1983).							X		X						X	X	
Required procedures	Number of procedures used to move an aircraft through sector airspace (Mogford et al., 1995).										X					X		
Requirements for longitudinal spacing and sequencing	Increased spacing requirements limit the allowed number of aircraft in sector due to fixed sector size (Mogford et al., 1995).										X					X		
Restricted areas, warning areas and military operating areas	Restricted areas limit airspace available for spacing and sequencing aircraft, like reducing sector size (Mogford et al., 1995).										X					X	X	
Retrospective verbalization	Participants reflect and verbalize about what is going on in a pre-recorded ATC situation. Used to identify cognitive structures and decision-making strategies (Leplat & Hoc, 1981).								X							X		

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Scenario Complexity Rating	Subjective rating of scenario complexity including traffic mix, crossings, sector size, weather cells, etc (Magyarits & Kopardekar, 2000).										X				X	X	X	X
Scenario Fidelity Rating	Overall subjective rating of the experimental fidelity (Allendoerfer & Galushka, 1999).														X	X	X	X
Secondary Task	This technique requires operators to perform a primary task within its specified requirements, and to use any spare attention or capacity to perform a secondary task. The decrement in performance of the secondary task is operationally defined as a measure of workload. Secondary task uses in research vary widely (e.g., card sorting, mental mathematics, identification, lexical decision, tracking and problem solving) (Bergeron, 1968; Colle, Amell, Ewry, & Jenkins, 1988; Slocum, Williges, & Roscoe, 1971; Vidulich, 1989b).					X		X								X	X	
Sector size	Square mileage a sector occupies. The smaller the sector the greater the complexity and task load (Mogford et al., 1995).										X					X		
Shell for Performing Verbal Protocol Analysis	This automated tool has been used successfully to aid analysis of concurrent verbal protocols (Sanderson, 1990).								X							X		

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Simulation of equipment errors and/or failures - Simulation Conditions	Scenario variable where equipment failures test the participant's ability to work under degraded modes of operation (Paul, 1990)			X	X	X		X								X	X	
Situational Awareness - computational model	The model has 3 components: 1) situational elements, 2) context-sensitive nodes, and 3) a regulatory mechanism that assesses situational elements for all nodes (Shively, Brickner, & Silbiger, 1997).					X											X	
Situational Awareness Global Assessment Technique	SAGAT can be used to focus on any tasks within situation assessment, e.g., acquiring the elements of a current situation; integrating relevant elements into a picture; and evaluating it (Endsley, 1988; 1990; 1994; 1996).					X										X	X	X
Situational Awareness Linked Instances	The method documents all operator actions throughout the session. Graphs showing an operator's transition from closed to open loop performance reveal changes in performance in complex systems (Muniz, Salas, Stout, & Bowers, 1998).								X							X	X	X

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Situational Awareness Linked Instances Adapted to Novel Tasks	The SALIANT method has five phases: 1) identify team SA behaviors, 2) develop scenarios, 3) define acceptable responses, 4) write a script, and 5) create a structured form with columns for scenario and responses (Muniz, Stout, Bowers, & Salas, 1998).					X										X	X	
Situational Awareness Rating Technique	The technique assumes that SA comprises 3 aspects of operator's task: (1) attentional resources, (2) demands on those resources, and (3) understanding of situation (Taylor, 1989; Taylor, Selcon, & Swinden, 1995; Selcon & Taylor, 1989).					X											X	
Situational Awareness Subjective Workload Dominance	SA SWORD uses judgment matrices to assess SA (Fracker, 1989; Fracker & Davis, 1991).					X										X	X	
Situational Awareness Supervisory Rating Form	Developed to measure SA in X-15 pilots. The form's 31 items range from general traits to tactical employment (Carretta, Perry, & Ree, 1996).					X										X	X	X
Slant Range Miss Distance -measure of aircraft separation	The shortest distance between two aircraft in conflict, measured by a straight line between the aircrafts' centers (Paul, 1990).		X											X		X	X	

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Snapshot of aircraft within a user-specified distance or time frame surrounding an event.	Offers ability to go back into the data and extract events surrounding a specific incident (such as an intentional blunder) (Ozmore & Morrow, 1996; Wing et al., in press).				X				X							X	X	
Spatial aspects of the controller's mental model	Controllers shown static air traffic scenarios involving aircraft pairs are asked to draw on paper the predicted relationship of the aircraft at the point of least separation (Lafon-Milon, 1981).								X							X	X	X
Speed	Frequency of speed clearances issued during a run (CRM, 1989; Ozmore & Morrow, 1996; Sollenberger, Stein & Gromelski, 1997).				X						X					X		
Speed adjustments	Frequency of FC initiated airspeed adjustments (FAA & NASA, 2001).							X				X		X		X	X	
Standard conflict cumulative durations variable	3 miles lateral and 1,000 foot vertical (Buckley et al., 1983; Boone et al., 1983; Paul, 1990).		X											X		X	X	
Standard conflict duration variable	5 miles lateral and 1,000 foot vertical (> FL290 = 2000 foot vertical) (Buckley et al., 1983; Boone et al., 1983; Paul, 1990).		X											X		X	X	
Standard conflict en route variable	5 miles lateral and 1,000 foot vertical (> FL290 = 2000 foot vertical) (Buckley et al., 1983; Boone et al., 1983; Paul, 1990).		X											X		X	X	

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Standard conflict terminal variable	3 miles lateral and 1,000 foot vertical (Buckley et al., 1983; Boone et al., 1983; Paul, 1990).		X											X		X	X	
Strip bay flight strip management	Measure of how well participant manages flight strips (Galushka et al., 1995).							X				X				X		
Structured Interviews	Participants were asked questions about their action priorities under normal and heavy workloads. Actions rated included scanning plan view display, sequencing traffic, calling and coordinating, and determining cross points (Redi & Nygren, 1988).				X											X	X	X
Subjective Performance Prediction	Subjective judgments by subject matter experts (SME) to predict operator performance. Judgments may be made about system design alternatives, procedural alternatives etc. (Sollenberger, Stein & Gromelski, 1997).								X							X	X	X
Subjective Workload Assessment Technique	The SWAT combines ratings of three different scales to produce an interval scale of mental workload including: time load, mental effort, and psychological stress (Bateman & Thompson, 1986; Derrick, 1983; Kilmer et al., 1988).							X								X	X	X

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Subjective Workload Dominance	Allows subjects to make pair-wise comparative ratings of competing design concepts along a continuum showing the degree to which one concept entails more or less workload (Vidulich, 1989a).							X								X	X	X
Subsequent conflicts	Frequency and duration of conflicts that result from a resolution of a previous conflict (Wing et al., in press).				X							X					X	
System Effectiveness Measures	The SEM set measures factors associated with system safety and efficiency: confliction, occupancy, communication, and delay (Buckley et al., 1983; Empson, 1987; Hopkin, 1980; Means et al., 1988).								X							X	X	
System Usability	Measure of the usability of the overall system (Allendoerfer & Galushka 1999; American National Standards Institute, 1993).					X	X	X				X				X	X	X
Task Analysis Workload	Technique decomposes missions into phases, segments, functions, and tasks. Subjective matter experts rate each task workload on a scale of 1 to 7. Task workloads are estimated for each point along a scenario timeline (Hamilton, Bierbaum & Fulford, 1991).						X	X								X	X	X

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Task Load	Task load is time required to perform a task divided by the time available to perform it. Values < 1 indicated excessive task load (Gunning & Manning, 1980).						X	X								X	X	X
Taskload per aircraft	Number of tasks or operations performed per aircraft (Galushka et al., 1995).							X				X				X		
Temporal Awareness	Ability of the operator to build a representation of the situation including recent past and near future. Measured as the number of temporal and ordering errors in a production line task, number of periods in which temporal constraints are adhered to, and temporal landmarks reported by operator to perform task (Grosjean & Terrier, 1999).			X	X	X										X	X	X
Time back on preferred trajectory	Delta path length and time to return to preferred route after constraints force a deviation (Wing et al., in press).				X							X				X	X	X
Timed performance of functions	Measures of task times to complete various functions (Adelman, Cohen, Bresnick, Chinnis, & Laskey, 1993; Frankish & Noyes, 1990).						X	X				X				X	X	X
Timeliness of ATC FF cancellation	Extent to which a controller allows pilot self-separation in a shared authority environment (FAA & NASA, 2001).				X	X		X				X	X	X		X	X	

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Traffic characteristics	Fidelity of the simulated traffic as representative of the real world (Galushka et al., 1995).										X				X	X	X	X
Trial plans	Number of trial plans (URET) by time and sector (FAA & NASA, 2001).				X	X		X					X			X		
Trust - ATC trust in pilot decision making role for separation	Metric needs further definition.			X								X				X	X	
Trust – ATC/FC trust in automation	Metric needs further definition.			X								X	X			X	X	
Trust- AOC in pilot/ATC decision making for optimal routes	Metric needs further definition.			X				X									X	X
Unified Tri-services Cognitive Performance Assessment Battery	Consists of 25 tests (e.g. grammatical reasoning, continuous recognition, visual scanning) selected based on: (1) use in one or more DoD laboratory, (2) proven validity, (3) relevance, and (4) sensitivity to hostile environments and sustained operations (Perez, Masline, Ramsey & Urban, 1987).								X							X		
Uninterrupted dwell points alternating between two ATC display objects	Sequentially examines relative positions between aircraft, and aircraft to geographical points on the display (Credeur et al., 1993).								X							X		

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Unstructured Group Discussion	Unstructured group discussion asks questions that participants received prior to discussions. This technique explores the concept of the controller's picture or SA (Mogford, Murphy, Yastrop, Guttman & Roske-Hofstrand, 1993).				X											X	X	X
Use of halo (J Ring)	Number of times J-ring or halo was used during experimental run (Galushka, Frederick, Mogford & Krois, 1995).							X						X		X		
Use of unusually high traffic rates to maximize pressure on the controllers - Simulation Conditions	Scenario variable where unusually high traffic loads present a stress test to the controller (Paul, 1990).							X			X	X			X	X		
User specifiable conflict variable	User specifiable conflict criteria for lateral and vertical separation (Buckley et al., 1983; Boone et al., 1983; Paul, 1990).		X											X		X	X	
User specifiable cumulative durations variable	User specifiable conflict criteria for lateral and vertical separation (Buckley et al., 1983; Boone et al., 1983; Paul, 1990).		X											X		X	X	
User specifiable terminal variable	User specifiable conflict criteria for lateral and vertical separation (Buckley et al., 1983; Boone et al., 1983; Paul, 1990).		X											X		X	X	

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Utilization	Probability of the operator being in a busy status. Accounts for arrival time of work in a queue and service time on that work (Buckley et al., 1983; Boone et al., 1983; Paul, 1990).							X								X	X	X
Vector lines	Number of times vector lines were used in experimental run (Nieva, Fleischman & Rieck, 1985).				X									X		X		
Verbal Protocol Analysis	VPA goal is to map unfolding of incidents as a scenario is completed. Types include think-aloud protocols, retrospective verbal reports and retrospective verbal reports (Selcon & Taylor, 1991).								X							X	X	X
Vertical distance between aircraft (feet)	Vertical component of slant range (in feet) (Paul, 1990).		X											X		X	X	
Vertical separation (feet)	Vertical separation of an aircraft pair in conflict (in feet) (Buckley et al., 1983).		X													X	X	
Voice duration -	Total duration of communications during run (FAA & NASA, 2001; Ozmore & Morrow, 1996; Porterfield, 1997).	X						X				X				X	X	
Voice frequency -	Frequency of push-to-talks accumulated during run (FAA & NASA, 2001; Ozmore & Morrow, 1996).	X						X		X		X				X	X	

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Weather	Weather produces complexity by limiting airspace available for maneuvering, blocking airways, and limiting altitudes available for vertical spacing (Mogford et al., 1995).									X	X	X	X			X	X	
Within-sector coordination (R&D teamwork)	Measure of taskload generated by the coordination between radar and data controllers (FAA & NASA 2001; Galushka et al., 1995; Sollenberger, Stein, & Gromelski, 1997).	X						X								X		
Workload Compensation Interference / Technical Effectiveness	The WCI/TE rating scale requires participants to rank 16 matrix cells and then rate specific tasks. Ratings are converted by conjoint scaling techniques to values of 0 to 100 (O'Donnell & Eggemeier, 1986; Wierwille, Casali, Connor, & Rahimi, 1985).							X								X	X	X
Workload guidelines	Subjective or objective estimates of workload as a set of task demands, effort, and activity (Chiles, Jennings, & Alluisi, 1979; Gartner & Murphy, 1979; Gopher, 1983; Hedge, Borman, Hanson, Carter, & Nelson, 1993; Hopkin, 1979; Leighbody, Beck, & Amato, 1992; McKenzie, Buckley, & Sarlanis, 1966; Meshkati, Hancock, & Rahimi, 1990; Rolfe, 1971; Wierwille & Eggemeier, 1993; Wierwille, Williges, & Schiflet, 1979).							X								X	X	X

METRIC	DEFINITION/DESCRIPTION	COMMUNICATIONS	CONFLICT	ERROR	DECISION MAKING	SITUATIONAL AWARENESS	USABILITY	WORKLOAD	OTHER	CAPACITY	COMPLEXITY	EFFICIENCY	FLEXIBILITY	SAFETY	SIMULATION FIDELITY	ATSP	FC	AOC
Zachery/Zaklad Cognitive Analysis	Requires both operational SME and cognitive scientist[s] to identify operator strategies for performing all tasks listed in a detailed cognitive mission task analysis. A second group of SME then rates, using 13 subscales, workload associated with performing each task (Zaklad, Zachary, & Davis, 1987).							X								X	X	X

A.2 LITERATURE REVIEW SUMMARY

The literature review presents human factors metrics that are appropriate for the ATM domain. Common metrics identifying the impact of the distribution of authority that can be transferred between the ATSP, AOC and FC need to be identified.

Most of the studies reviewed are related to ATSP and FC operations. The literature review indicates that the AOC metrics are not well understood. Further metric definition is necessary in the area of shared separation authority, decision-making, trust, system flexibility, and predicted error. In addition, research is required to examine the interrelationships of human performance as separation authority is distributed between the ATSP, AOC, and FC.

The researchers will continue to expand this database during the course of DAG-TM evaluations. Researchers will work in conjunction with CE leads and SME to define metrics for DAG-TM specific concepts that are feasible and beneficial. Metrics that overarch ATSP, FC, and AOC domains will be identified and guidance provided for the application to CE concepts.